2018

Modern approaches to sepsis - evolving definitions, clinician roles, and AI-based diagnostic aids

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

MODERN APPROACHES TO SEPSIS – EVOLVING DEFINITIONS, CLINICIAN ROLES AND AI-BASED DIAGNOSTIC AIDS

by

JASON MATHIAS

B.A., University of Southern California, 2010
M.F.A., University of Southern California, 2013

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2018

Approved by

First Reader _____

William E. Baker, M.D.
Clinical Associate Professor of Emergency Medicine

Second Reader _____

David Flynn, MS (LIS)
Associate Professor of Library & Information Management Education

DEDICATION

I would like to dedicate this work to those who have opened my eyes to the possibility of truly enjoying my career. My truest friends, Loan Verneau, Alex Beachum, Kedar Reddy, Casey Sakima, who have been a moral, intellectual, and emotional compass throughout the last 10 years. Herman and Blanche Mathias, for being wiser than me, and for letting me learn through experience. You all have always been on my team, even when I wasn't.

ACKNOWLEDGMENTS

I would like to thank Elena David, Josh Green, Michelle Leisner, Lauren Johnson, Oliver Hayworth, and Caitlin Long for providing me the best experience of living in Boston I could ask for. For being excellent examples of teaching, I would like to thank Louis J. Toth and Chris Lim, as well as Fernando Garcia-Diaz, for holding a high standard for my learning and making sure I could answer the toughest questions.

I would also like to thank David Flynn, for showing me how to be smart, and clever, in the pursuit of academic research, Dr. William Baker, for his insight on the practical realities behind the academic papers, Nitun “Bell” Varongchayakul, for help in the early stages of thesis, and Nicholas Cilfone for answering my countless questions about how machines think.

MODERN APPROACHES TO SEPSIS – EVOLVING DEFINITIONS, CLINICIAN ROLES AND AI-BASED DIAGNOSTIC AIDS

JASON MATHIAS

ABSTRACT

Sepsis is an ongoing concern in critical care. It is hard to quickly detect, and rapid deterioration of a patient into septic shock causes death in around 30% - 50% of patients, while survivors may live with organ damage and shorter lifespans. Traditional methods of detection require long laboratory tests and clinician vigilance, which put a strain on hospital resources.

New advances in machine learning offer an alternative – using algorithmic analysis in real-time to watch for a deteriorating patient state. The use of readily available data – heart rate, respiratory rate – combined with electronic medical records and fast laboratory tests presents an opportunity for early detection of sepsis, which can potentially make great strides in minimizing damage to patients.

A variety of algorithmic methods have been proposed by researchers, and research so far has been promising. Algorithms inretrospective studies have performed equal or better to standard protocols such as SIRS or SOFA. Some promising research even presents the opportunity to approach sepsis diagnosis and treatment in an entirely new manner. At the present stage, however, the field is at too early a stage for use in a clinical environment. This review intends to review some prominent types of machine learning algorithms, as well as discuss current concerns regarding machine learning-based detection support systems (ML-DSS).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BU ..... Boston University

CRP ..... C-Reactive Protein

HMM ..... Hidden Markov Model

MAP ..... Mean Arterial Pressure

MFA ..... Master's of Fine Arts

ML-DSS ..... Machine Learning-Based Decision Support System

NLP ..... Natural Language Processing

PCT ..... Procalcitonin

SIRS ..... Systematic Inflammatory Response Syndrome

SOFA/qSOFA ..... Sepsis-Related Organ Failure Assessment/Quick SOFA

WBC ..... White Blood Cell

INTRODUCTION

Sepsis is a prevalent concern in critical care scenarios, and a leading cause of mortality for hospitalized patients. Tromp et al. estimates that 2% of all hospitalized patients are diagnosed with severe sepsis or septic shock.[1,2] Later stages of sepsis have high morbidity, and early detection is very important for chances of good outcome. While specific definitions of sepsis are controversial, the mechanism of sepsis occurs from infection. A pathogen infects the bloodstream, triggering a systemic inflammatory response. If the response is large enough to become dysregulated, that can lead to organ failure in patients[3]. Upon onset of sepsis in an in-hospital setting, there is a 20-30% rate of mortality. Patients that survive to discharge often have increased risk for death in the following months and years, and can endure mood disorders and impaired physical or neurocognitive functioning[4].

Unfortunately, sepsis detection is a complicated process. There is still no single unifying biomarker that clearly demarcates sepsis[5] – rather, an evolving set of criteria, combining biomarkers and physiological symptoms, is used to diagnose the condition as rapidly as possible to allow for speedy administration of treatment. Some of these criteria look for consequences of organ damage brought about by inflammation and ischemia[6], while others are assessments of visible symptoms such as changes in body temperature or respiratory rate. An example of this set of criteria is the Sepsis-III definition, which takes a series of physiologic parameters (heart rate, respiratory rate, etc.) and plugs them into an equation, a high enough score indicating a sepsis condition[7].

Making sepsis diagnosis more complicated is the high human resource requirement – A clinician has to take a series of vital signs, and using these values, calculate a 'score' for whether the patient shows symptoms of sepsis. Due to the complexity of the clinical setting, this can be a strain on already strained resources. Given this condition, there has been an interest in using computer technology as a diagnostic aid. There have been successful attempts to automatically sense when vital signs correspond to the guidelines of a protocol like Sepsis-III, but there is a limit to this usefulness – these systems will only alert at the "sicker stages" of onset of sepsis, rather than when a patient is in the early stages.

Sepsis has a high mortality rate, but even sepsis survivors face additional morbidity. Re-hospitalization occurs within a year for the majority of survivors,[8] and for patients with sepsis and organ dysfunction, there is a 74% chance of mortality within 5 years[9]. While having a computer process alert clinicians of sepsis is useful (as it could potentially decrease sepsis severity and therefore lower mortality rates), it would be more ideal to have a system that could predict sepsis before it presents in the patient, to prevent major organ damage and further increase patient morbidity.

With the recent rise in machine learning, there has been an interest in applying advanced algorithms to create a Machine Learning-Based Diagnostic Support System (ML-DSS) from biomarkers and readily-available Electronic Health Record (EHR) recordings. In theory, a properly trained algorithm could give a confidence level of how likely it is that the patient is becoming septic, before manual scores would even show a sepsis diagnosis. A few promising retrospective studies have been conducted, as well as

some live clinical studies. While the initial results are promising, more research should be conducted before machine learning algorithms are robust enough – and integrated into clinician workflow well enough – to be regularly used in an Emergency Department or Intensive Care Unit.

The purpose of this thesis is for clinicians with minimal computer programming knowledge, and computer science professionals with minimal clinical experience. It is intended to be a primer on the current state of sepsis, diagnosis and treatment paradigms for sepsis, an introduction to machine learning algorithms, and an exploration of issues related to machine learning implementation.

One important limitation of note is that definitions for sepsis recently changed in 2016, and as such, academic literature may not be comparable when discussing the same subject. Even literature after 2016 may refer to the older SIRS-based definition of sepsis. Until 2016, Sepsis was defined as a pathogenic infection in the bloodstream, combined with the host response (SIRS)[10]. In 2016, the condition was redefined as a "life-threatening organ dysfunction caused by a dysregulated host response to infection."[7] In both definitions, the presence of a pathogen in the blood causes an immune response, and since the bloodstream allows for fast movement throughought the body, the potential for pathogen damage (and immune response) becomes pervasive throughout the body. Prior to 2016, the presence of organ dysfunction was considered 'severe sepsis', and where possible, this article uses the definition of 'severe sepsis' to minimize confusion.

SEPSIS MECHANISMS

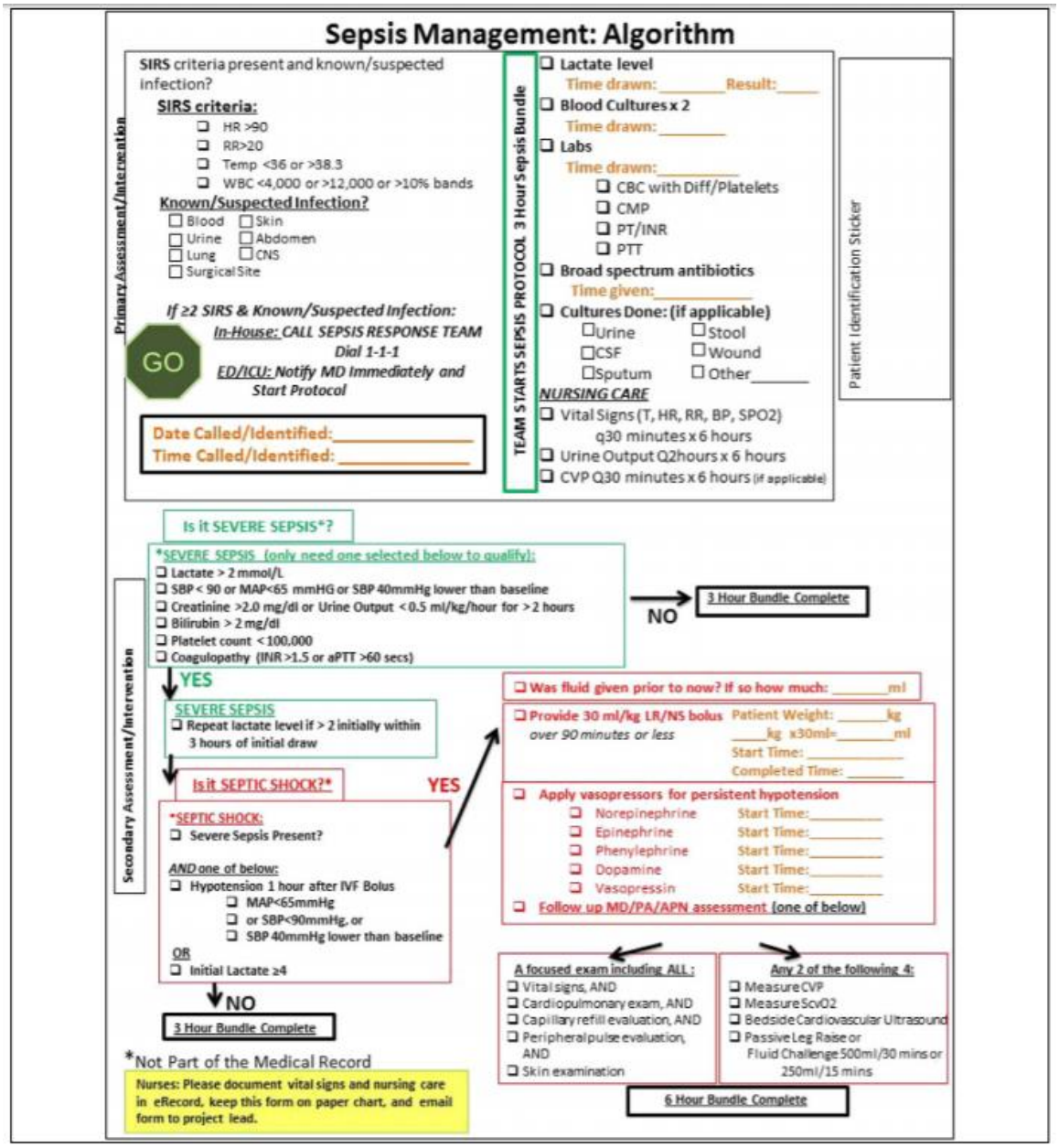
While the complete mechanisms for symptoms of sepsis are not fully understood, part of the mechanism is a bodily response that causes low oxygen perfusion to necessary tissues, which leads to impaired organ function[11]. Cytokine release at the presence of infection causes an immune response[12], as well as an increase in coagulative response, as well as inhibition of fibrinolysis.[13] This is normally intended to isolate a pathogen[14] - if an infection is limited to a small, separated area, low oxygen perfusion works to induce cell death, which starves the infection of a nutrient supply. If this system is enacted throughout the body, however, the effect is not pathogen isolation, but potentially widespread damage. If not treated, this can lead to organ failure, and the onset of septic shock, which is defined as low blood pressure brought on by sepsis[15].

Since the underlying mechanism for sepsis is an infection in the bloodstream, the gold standard for sepsis diagnosis remains the presence of identifiable bacteria in the bloodstream[16]. However, pathogens only present in blood cultures only about 30% of the time[17], so looking for proof of infection can be inconclusive. One issue of complexity is that there are multiple pathways that can lead to sepsis – for some patients there is an immune-mediated response, whereas in others there may be an immune suppression or an accelerated lymphatic apoptosis[15]. As such, sepsis can have a variety of symptoms and markers, but the absence or presence of a specific biomarker may not actually be indicative of sepsis.

ISSUES OF SEPSIS DIAGNOSIS AND TREATMENT

There is a substantial increase in mortality from sepsis if the patient is diagnosed with sepsis while in a hospital setting. If sepsis patients are separated by whether they Present on Admission (POA) or are diagnosed while in the hospital setting (NPOA, or "Not Present on Admission"), the POA have about a 15% mortality rate, while NPOA patients have a significantly higher mortality rate of about 35%[18]. To some extent, this is because sepsis is difficult to diagnose – SIRS may be a symptom of an impending myocardial infarction rather than a bloodstream infection, which would require different treatments.[18]

For rapid diagnosis, clinicians will often calculate a score, either the Systemic Inflammatory Response Syndrome (SIRS)[19] score or the Quick Systemic Organ Failure Assessment (qSOFA)[7] score. Manual paper-based aids are sometimes used for nurses to calculate a diagnostic SIRS or SOFA score, which can be time-intensive. An example is shown as figure 1.[20]



FIGURE

Sepsis management algorithm tool. *APN*, Advanced practice nurse; *aPTT*, activated partial thromboplastin time; *BP*, blood pressure; *CBC*, complete blood cell count; *CMP*, comprehensive metabolic panel; *CSF*, cerebrospinal fluid; *CVP*, central venous pressure; *HR*, heart rate; *ICU*, intensive care unit; *INR*, international normalized ratio; *IVF*, intravenous fluid; *LR*, lactated Ringer's solution; *MAP*, mean arterial pressure; *MD*, physician; *NS*, normal saline solution; *PA*, physician assistant; *PT*, prothrombin time; *PTT*, partial thromboplastin time; *Q*, every; *RR*, respiratory rate; *SBP*, systolic blood pressure; $Scvo_2$, central venous oxygen saturation; *SIRS*, systemic inflammatory response syndrome; Spo_2, oxygen saturation; *T*, temperature; *WBC*, white blood cell count.

Figure 1. Example of an algorithmic guideline for nurses in Tedesco, et al. Nurses were encouraged to fill out the tool by hand, as a way to assess the severity of sepsis in a patient.

Human error and logistical difficulty add to the issues of sepsis treatment. The Surviving Sepsis Campaign recommends a series of four steps to be completed within 3 hours of admission of a patient with severe sepsis: "(1) measure the serum lactate level, (2) obtain blood cultures before antibiotic initiation, (3) administer broadspectrum antibiotics, and (4) infuse 30 mL/kg of an intravenous (IV) crystalloid solution in patients with hypotension or a lactate level of 4 mmol/L or greater" (Bruce et al)[21]. Steps that require multidisciplinary coordination (such as antibiotic and fluid administration) are often not performed in that timeframe – and since one of the most important factors for prognosis is timely treatment, any delays can increase mortality rates. In a study of interprofessional coordination with regards to sepsis treatment administration, when compliance with the three-hour window improved, mortality from sepsis decreased significantly, even in a time of increased sepsis diagnosis.[20]

BIOMARKERS AS A TOOL OF DIAGNOSIS

To complicate matters, many of the symptoms of sepsis can present in other conditions such as pancreatitis[22]. Early signs of sepsis include symptoms such as fever and tachycardia, which, while indicative of sepsis, are also indicative of other non-infection-based systemic inflammatory response syndromes.[23] Symptoms could also be indicative of infections that have not gotten into the bloodstream - In one study, automated screening tools were implemented to flag any patients that met criteria for

SIRS. When the results were analyzed, only 44% of the patients ended up having sepsis, though a large percentage of the non-sepsis conditions were some sort of infection.[24]

Since it is important to confirm the presence of pathogens, laboratory biomarkers are often used for diagnosis. 178 biomarkers have been proposed for sepsis diagnosis[25], though it is important to note that no single biomarker exists that is specific only to sepsis. One reason so many biomarkers have been proposed is that sepsis has a complex pathophysiology – in addition to inflammation, coagulation, complement system activation and apoptosis are all part of sepsis pathophysiology, so markers have been proposed related to these, and other, steps in the process.[25]

A few promising biomarkers stand out – Procalcitonin, Hyperlactatemia, and C-Reactive Protein. Some of these biomarkers stand out for their specificity and sensitivity, while others do so because of their availability in the hospital setting.

Procalcitonin[23] has been proposed as a good potential biomarker, though its presence alone cannot indicate sepsis. Procalcitonin is virtually undetectable in healthy individuals, but as it is a prohormone for calcitonin, which is released as a response to bacterial infection, Procalcitonin levels can be used to measure bacterial infection.[26] Procalcitonin, is not highly specific or highly sensitive to sepsis[25] (increased levels are present in other inflammatory responses such as trauma[27] and major surgery[28]), but procalcitonin nevertheless has seen extensive use as a diagnostic tool.

CRP is primarily produced by hepatocytes as a response to IL-6,[29] but in comparison to an IL-6 laboratory test, CRP is more readily available. The presence of CRP is correlated with organ failure, to the point that persistently high levels of CRP

correlate with poor prognosis.[30] However, CRP is also not specific to sepsis, and is increased in other inflammatory diseases.[10]

Hyperlactatemia[6] is another promising biomarker – lack of oxygenation causes tissues to produce lactic acid as a byproduct of anaerobic metabolism, so high levels of lactic acid can be monitored to monitor for the effects of SIRS.

In the case of neonates, cytokines and cell-surface molecules tend to be over-expressed, so diagnostic tools sensitive to these signals provide a promising option for neonatal sepsis diagnosis[16]. Two markers that are promising for neonates are IL-12, and Interferon-produced Protein 10 (IP-10).[25]

In a review of biomarker literature, Pierrakos et al. suggest that perhaps biomarkers are more effective for *ruling out* sepsis, rather than for diagnosing.[25] Specifically, PCT has a high negative predictive value for sepsis – that is, if levels of PCT are not high, it is unlikely that the patient has sepsis.[31]

Other markers, arising from various effects of sepsis, are presented in Table 1. For a more complete picture, Pierrakos, et al., in their review of biomarkers, present a series of more than 170 biomarkers, categorized based on aspects of sepsis they are designed to monitor.[25]

Table 1. Diagnostic value and limitations of biomarkers to separate infectious from non-infectious causes of inflammation

Biomarker	Source	Sens.	Spec.	AUC	LR^+	LR^-	Limitations
C-reactive protein[21]	Metaanalysis (n = 1386)	0.75	0.67	–	2.43	0.42	Slow kinetic, independent of infection severity, increased in many inflammatory diseases
Procalcitonin[35]	Metaanalysis (n = 3244)	0.77	0.79	0.89	4.0	0.29	Increased in various non-infectious causes of SIRS (i.e., cardiac arrest, severe trauma)
Interleukin-6[57]	Cohort study (n = 327)	0.82	0.75	0.86	–	–	Limited data, conflicting results
sTREM-1[78]	Metaanalysis (n = 1795)	0.79	0.80	0.87	4.0	0.26	Present in inflammatory disease without infection
LBP[57]	Cohort study (n = 327)	0.57	0.85	0.73	–	–	Non-specific marker of inflammation
suPAR[98]	Cohort study (n = 273)	–	–	0.62	–	–	Limited data; low diagnostic value for sepsis

Data give sensitivity (sens.), specificity (spec.), area under the curve (AUC) from receiver operating characteristics, positive (LR^+) and negative (LR^-) likelihood ratios of a biomarker for differentiation of infectious vs. non-infectious causes of inflammation. LBP, lipopolysaccharide binding protein; suPAR, soluble urokinase plasminogen activator receptor; sTREM 1, soluble triggering receptor expressed on myeloid cells 1.

Table 1. Markers used to differentiate sepsis, along with limitations. From Bloos, Reinhart et al[10].

COSTS OF SEPSIS TO PUBLIC HEALTH

In 2013, Sepsis represented the most expensive condition treated in the United States.[32] At roughly 23.7 billion dollars, it represented more than 6% of the total costs of all hospitalizations. A large part of the cost stems from re-hospitalizations – in the state of California alone, between 2009 and 2011, re-hospitalization due to sepsis represented a $500 million cost annually, more than the re-hospitalization costs of acute myocardial infarction and congestive heart failure combined.[33]

Survivors of sepsis suffer additional morbidities. There is a five-fold increase in 5-year mortality for survivors of an index sepsis episode.[34] 60% of sepsis survivors are

re-hospitalized within a year, most often due to infection. Of those that are re-hospitalized, there is a 17% mortality rate.[8] For infection-related rehospitalizations, more than half of the cases were reported to be due to an unresolved or recurrent infection relating to the index sepsis event.[35]

Costs for both index sepsis cases and re-hospitalizations are high. In a retrospective analysis of costs between 1995 and 1998, the average cost of an index hospitalization for sepsis with organ dysfunction was about $27,000, with patients requiring an ICU visit averaging about $36,000[36]. (It is important to note that costs are highly variable. The standard deviation for a mean of $27,000 was $55,000, reflecting the variable nature of individual costs depending on patient conditions such as co-morbidities).

For re-admissions, costs increased by another $11,000 to $25,000, depending on whether the patient survived past 12 months.[36] A similar study conducted around 2011 data showed a similar cost of $25,000 for hospital readmissions, with the most common (22%) cause of readmission being sepsis, though it was not clear whether this was due to a recurrent sepsis or new infection.[37]

EVOLVING DEFINITIONS OF SEPSIS

The definition of a clinical diagnosis of sepsis is not universally agreed upon. While the gold standard of Sepsis remains the proof of infection in the bloodstream, the clinical presentations of the condition are varied and can be different depending on factors such as age and compounding factors such as alcohol use disorder.

In 1991, SIRS was defined by Bone et al. [22] as a clinical standard to diagnose Sepsis. Sepsis was defined as SIRS with a suspicion or finding of blood infection, which came to be known as the "Sepsis-1" definition.

SIRS is defined as a patient having two of the following four conditions: "tachycardia (heart rate >90 beats/min), tachypnea (respiratory rate >20 breaths/min), fever or hypothermia (temperature >38 or <36 °C), and leukocytosis, leukopenia, or bandemia (white blood cells >1,200/mm3, <4,000/mm3 or bandemia ≥10%)" (Marik et al)[19]. A diagnosis of sepsis requires proof or a suspicion of infection, as well. There are three states of sepsis within Sepsis-I – 'sepsis', 'severe sepsis' (organ dysfunction brought upon by sepsis), and 'septic shock' (low blood pressure brought on by sepsis).

One of the advantages of Sepsis-1 was that its diagnosis required largely external criteria – heart rate, respiratory rate, and temperature can all be readily assessed via a machine or clinician observation, leaving only white blood cell count (WBC) as a test that required laboratory examination. However, this was not without consequence - while SIRS became widely standardized for the diagnosis and treatment of sepsis, it has been criticized for a low specificity.[38] Non-sepsis conditions such as ischemia, tissue injury, and pancreatitis can present with SIRS, [22] and since the recommended treatment for each condition is not the administration of antibiotics, there was concern of potential misdiagnosis and treatment that does not improve the patient's condition. In addition, in one study, 13% of patients presented as negative for SIRS, yet still had sepsis.[39]

In 2016, an attempt to address this issue of low specificity was made by the Society of Critical Care Medicine (SCCM) and the European Society of Intensive Care

Medicine (ESICM). The new definition, known as Sepsis-III, largely abandons SIRS as a criteria[19]. Sepsis-III redefines sepsis as "life-threatening organ dysfunction caused by dysregulated host response to infection"[7]. Before this definition, 'Sepsis' was defined as the presence of bloodstream infection and SIRS, while the presence of organ dysfunction was considered 'severe sepsis'. With the new Sepsis-III definition, 'severe sepsis' has largely become renamed to 'sepsis', and infection without organ dysfunction is sometimes referred to as 'pre-sepsis.'[40]

In re-casting sepsis as a condition of organ dysfunction (as opposed to inflammatory response), SIRS detection was replaced with criteria for the detection and diagnosis of organ dysfunction. The proposed measure was the Sequential Organ Failure Assessment (SOFA) score, an assessment found to be more predictive for in-hospital mortality than the SIRS criteria. SOFA looks at six major categories, rather than four, and assigns point values from 0 to 4. Higher scores indicate higher mortality rates, and the direction of change in SOFA points indicates the direction of patient prognosis. In other words, if the SOFA score increases, the patient is going into a worse condition. Increases in SOFA score of more than 2 points to potential organ dysfunction.[41] A table of the SOFA scoring sheet and score analysis is presented in Table 2.

Note that the SOFA score uses multiple laboratory tests (Platelet, Bilirubin concentration) and also a measure of mental consciousness impairment (Glasgow Coma Score)[42]. It also monitors multiple organ systems. A simpler analysis, the quick Sequential Organ Failure Assessment (qSOFA) presents a faster alternative, designed for rapid diagnosis within an intensive care unit. The qSOFA uses only three measures – low

blood pressure (low blood pressure (SBP≤100 mmHg), high respiratory rate (≥22 breaths per min), or altered mentation (Glasgow coma scale<15).[43] Each of these measures are accessible without bloodwork, allowing for a relatively rapid analysis which can then be followed up with a more comprehensive SOFA analysis if needed.

Example SOFA scoring sheet;

Sequential Organ Failure Assessment		0	1	2	3	4	Score
Respiratory	PaO2/FiO2, mmHg (kPa)	≥400 (53.3)	399-300 (40-53.3)	299-200 (26.67-40.0)	199-100 (13.33-26.67) AND	<100 (<13.33) AND	
			350				**1**
	Respiratory support (yes/no)				Yes	Yes	
Coagulation	Platelet (x 103/microliter)	≥150	<150	<100	<50	<20	
				95			**2**
Hepatic	Bilirubin mg/dL (mmole/L)	<1.2 (<20)	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	>12 (>204)	
			1.5				**1**
Circulatory	Mean arterial pressure (mm Hg)	≥70	<70				
			68				**1**
	Dopamine dose (mg/kg/min)			≤ 5 **OR**	> 5 **OR**	> 15 **OR**	
	Epinephrine dose (mg/kg/min)				< 0.1 **OR**	> 0.1 **OR**	
	Norepinephrine dose (mg/kg/min)				< 0.1	> 0.1	
	Dobutamine (yes/no)			Yes			
Neurological	Glascow coma score	15	13-14	10-12	6-9	< 6	
			13				**1**
Renal	Creatinine level mg/dL, (mmol/L)	<1.2 (<110)	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-440) **OR**	>5 (>440) **OR**	
			1.5				**1**
	Total urine output (ml/d)				201-500	< 200	
						Total SOFA score	**7**

Maximum SOFA Score	Mortality (%)
0-6	< 10
7-9	15-20
10-12	40-50
13-14	50-60
15	>80
15-29	>90
Vincent JL, de Mendonça A, Cantraine F, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. Crit Care Med. 1998;26(11):1793-800.	

SOFA Score Trend (first 48 hr)	Mortality (%)
Increasing	>50
Unchanging	27-35
Decreasing	<27
Ferreira FL, Bota DP, Bross A, et al. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA. 2001;286(14):1754-8.	

Table 2. SOFA score categories and criteria. Note that increased score trends indicate significant chances of mortality. Table from Chamberlain, NR[36]

While the new Sepsis-III definition has generally been accepted in the hospital setting, it is not without criticism. One criticism stems from the loss of a definition for bloodborne infection without organ damage. Before Sepsis-III, any infection and dysregulated immune response would be considered sepsis, regardless of organ dysfunction. The change in definition has been seen as potentially precluding treatment for the early stages of infection[40], which could allow for better patient outcomes. In at least one study, qSOFA was found to be poorly sensitive for organ dysfunction,[44] and argued that the SIRS score conveys "more prognostic and clinical information",[44] even if SOFA presents a comparable analysis of mortality rates. Consensus has not been reached on the preference of SIRS or SOFA/qSOFA.

TREATMENT OF SEPSIS

For proper treatment of sepsis, early detection is critical. The longer a pathogen is in a bloodstream, the stronger the immune response will be, and the stronger symptoms such as organ dysfunction will be. As organ dysfunction increases, the chance for secondary complications increases, and mortality rates increase. At some point, conditions become bad enough to cause low blood pressure, known as septic shock. Once a patient has septic shock, every hour delay increases the chance of mortality by 8%.[45]

Once a diagnosis of sepsis, severe sepsis, or septic shock is made, treatment options focus on detecting the source of the infection, minimizing damage to the body, and eliminating pathogens. Guidelines have been established by the Surviving Sepsis Campaign which recommend a series of procedures.[11] Since there is likely a bacteria

causing sepsis, the first step is to understand what bacteria that is, so two blood cultures are taken, aerobic and anaerobic. Identifying the bacteria allows for more targeted administration of antibiotics, and pathogens tend to depend on the original infection site.[46] For instance, *Staphylococcus aureus* and coagulase negative staphylococci are the most causative organisms for meningitis and pneumonia[46], and *Escherichia coli* is the most prevalent cause of urinary tract infection-related sepsis.[47]

This, however, should only be done if it does not interfere with administration of antibiotics within an hour of diagnosis – if it does, the antibiotics take precedence.[11] The antibiotics should be broad spectrum at this point, or an antibiotic that works on both gram-positive and gram-negative bacteria. Considering the presence of drug resistant bacteria, some possible supplemental drug administration must be considered. For instance, if there is a risk of MRSA in the area, drugs such as vancomycin or teicoplanin are suggested supplements.[11]

Next, if the patient has hypoperfusion, the Surviving Sepsis campaign guidelines suggest administration of IV crystalline fluid, 30 ml/kg, within 3 hours. Finally, a lactate assessment should be regularly done, and the goal should be to normalize lactate levels, which implies proper tissue perfusion.[11]

COMPUTER-BASED DETECTION SYSTEMS

As early detection and treatment of sepsis dramatically affects prognosis[48], there has been an interest using computational systems to monitor patients without the need for a clinician. In theory, a computer system could actively monitor a patient's vital signs for

symptoms characteristic of early sepsis, and raise a warning to clinicians if the right criteria is met. At the most basic level, this would amount to a computer reading vital signs and laboratory data, tabulating a score based on Sepsis-3 or APACHE criteria, and alerting a clinician if the basic requirements for sepsis are met.

An example of this was conducted at Barnes-Jewish hospital in St. Louis[49]. Using an automated warning tool that required manual nurse entry of hemodynamic data and automated entry of laboratory data, the tool would generate and send an alert to staff if a patient registered as a potential case. All patients that registered an alert went on to have sepsis, severe sepsis, or septic shock within 48 hours of an alert being generated, and the early warning was reported to significantly increase the rate of interventions within 12 hours of a generated alert[49].

While this system is functional, it is limited to being, essentially, another sensor. It lacks the ability to predict worsening conditions, and will only alert clinicians at conditions that meet sepsis criteria, rather than when the patient is deteriorating into sepsis. At that point, there may already be damage to the body. Yet, detecting sepsis before it shows the established symptom criteria is tricky even for a clinician – Myocardial Infarctions, for instance, can sometimes also cause SIRS symptoms, so simply alerting clinicians every time SIRS is detected would not be an effective preventative measure. Simply setting notices for potential SIRS criteria could lead to alert fatigue, potentially neutralizing the benefits of an automated sepsis detection system.

MACHINE LEARNING : A PREDICTIVE, RATHER THAN REACTIVE, TOOL

As such, for a system to effectively alert clinicians to potential Sepsis situations, computer programs have to move past the idea of simply programming a checklist of symptoms, and move into smarter, Machine-Learning-based predictions.

Machine Learning is a broad umbrella term encompassing a variety of algorithmic approaches to combing data to make predictions. Broadly, machine learning takes a large set of data and uses statistical analysis to "learn" patterns in the data. Large volumes of data that are difficult for a clinician to assess and quickly analyze can be processed at significantly higher speeds by algorithmic analysis. For instance, patients in a critical-care units can be continuously monitored for heart rate and respiratory rate, and these data points can be analyzed for trends without human assistance. When combining this with electronic medical records, laboratory analysis of biomarkers or white blood cell (WBC) count, a machine learning algorithm has a potentially rich field of data to use in order to detect patterns in patients with sepsis. The specific diagnosis of sepsis requires some indication that the SIRS is infection-related, so the addition of biomarkers aid in telling a clinician that sepsis-specific treatments (as opposed to, say, cardiac arrest) are called for.

Machine Learning has recently seen a renaissance in the fields of statistics and computer science, and has since been applied to fields as separate as melanoma detection and language translation. The origins of what is currently called "machine learning" are rooted in statistical analysis, and trace back more than 30 years[50]. However, recent increases in both the volume of accessible data (including EHR data) and cheap

computational ability have allowed for a substantial increase in activity in the field.[51] In an abstract sense, machine learning represents a shift in program design – instead of programming a set of rules, the program designer creates an algorithm that can take an "unknown" value, use statistics to make 'guesses' about what that value means, and then check that 'guess' against the right answer. The system does not necessarily have a reference for what a specific value or condition means – it simply recognizes that certain values are statistically correlated with certain outcomes, and uses that analysis to make a prediction about newly presented information.

There are multiple styles within the study of machine learning, ranging from the relatively simple "logistic regression" to the very complex "deep learning". It is important to understand, however, that "more complex" doesn't necessarily mean "better" - in general, the more complex the machine learning algorithm, the more powerful the algorithm or the more data it can process, but the more difficult development and optimization becomes.

SEPSIS PREDICTION WITH MACHINE LEARNING

A number of machine learning experiments pertaining to sepsis prediction have been attempted. The studies range in algorithmic styles, from Deep Learning to Hidden Markov Models to more basic logistic regression. Almost all studies are retrospective analyses of EHR datasets, many from MIMIC-II and MIMIC-III. Retrospective analysis projects are interested in predicting sepsis that is occurring to a patient while in a hospital setting, so often they will remove data about patients entering with a diagnosis of sepsis,

or that is diagnosed with sepsis within a few hours of arrival.[52] Not all studies are retrospective – Algorithmic development company Dascena created a commercial product, InSight, as an ML-DSS tool, and performed a randomized clinical trial. The trial saw a 12.4% decrease in mortality when using Insight.[53]

The specific biomarkers and vital signs collected varies for each study. Shamim et al. used 65 different variables, ranging from blood pressure to fibrinogen, to the standard deviation of respiratory rate intervals[52]. Taneja et al. used six biomarkers, and found that specific biomarkers would be of peak importance at different stages of sepsis (for instance, nCD64 was more indicative of early sepsis, while high PCT levels were more indicative of latter stages of sepsis).[54] This was used to great effect for that study, as the algorithm could accurately suggest the severity of sepsis that the patient was in.

Interestingly, one of the first published studies was a prospective trial. In 2010, Tang et al separated patients coming into the Emergency Department with sepsis symptoms as either SIRS or Severe Sepsis. 28 total patients began, and 2 were excluded due to complications. It found that the algorithm was able to successfully classify SIRS versus severe sepsis, though they found issues with specificity.[55] They suggested a larger sample size and more comprehensive cardiovascular features as possible improvements for later studies.

In 2011, Ohno-Machado, editor of the Journal of the American Medical Informatics Association, called for an investigation into Natural Language Processing (NLP) for EHRs, noting that the unstructured handwritten notes provided in EHRs contain important information to use in research.[56] That same year, Sawyer et al.

published a report of a computerized non-machine learning alert system, which was based on "shock index (heart rate divided by systolic blood pressure), MAP, international normalized ratio, WBC count, hemoglobin, absolute neutrophil count, serum albumin, total bilirubin, and sodium.[49] The prospective clinical trial showed increased speed of intervention, antibiotic escalation, intravenous fluid administration, and oxygen therapy. It noted no change in hospital mortality or length of stay, however.[49]

Convertino et al. published a study of machine learning for diagnosing hemorragic shock, for both civilian and battlefield trauma. The study used lower body negative pressure as a model of hemorrhage until hemodynamic decompensation. The study was "96.5% effective in predicting the estimated amount of reduced central blood volume".[57] While not directly applicable to septic shock, the study does have some similarity and promising findings.

In 2014, Gultepe at al. focused on hyperlactatemia as a measure of potential septic mortality, using machine learning as a prediction tool. A retrospective study of 741 patients from the University of Southern California Davis were used, and their vital signs and white blood cell (WBC) counts were associated with sepsis occurrence and mortality. The intent was to use the algorithm to predict lactate level and mortality risk. Using only vital signs and WBC count, they were able to make predictions of lactate level with an accuracy of 0.99, and using mean arterial pressure, median absolute deviation of respiratory rate, and median lactate levels, they were able to predict mortality to an accuracy of 0.73 retrospectively.[58]

That same year, Mani et al. looked at neonatal sepsis diagnosis with machine learning, comparing nine ML algorithms to clinician diagnosis. The retrospective study looked at 299 neonatal patients from an 18 month period from January 2006. The study generated temporal variables from time-stamped measurements from the laboratory, in order to give a time-sensitivity to the processed information. Interestingly, the study looked at both culture-positive and culture-negative sepsis, as bacterial cultures don't always present a pathogen in cases of sepsis.[17] The results were significant – when including culture-negative sepsis, "the treatment sensitivity of all the nine ML algorithms and specificity of eight out of the nine ML algorithms tested exceeded that of the physician…When culture-negative sepsis was excluded both sensitivity and specificity exceeded that of the physician for all the ML algorithms. The top three predictive variables were the hematocrit or packed cell volume, chorioamnionitis and respiratory rate."[59]

Nguyen et al. published another study of a non-ML alert automated alert system, based on SIRS, in 2014. In the 795 automated sepsis alerts that were generated, there was a 44.7% prevalence of sepsis. In 300 randomly selected non-sepsis alerts, there was a 0% prevalence of sepsis. While a step in the right direction, the low accuracy suggests that a more complex algorithm for prediction was necessary.

Tsoukalas et al. (2015) noted the difficulty of making a framework for clinical decision support when data was incomplete, as in EHR records. Nevertheless, they used a larger dataset of 1492 patients, and used a form of the Markov Model (explained below) to derive an optimal policy for treatment. When that optimal policy was followed, there

was an increase in transition to better states (25.9% vs. 12.9%) and a decrease of transition to worse states (33.7% vs. 51.2%).[60]

In 2016, Dascena, a diagnostic algorithm company specializing in ML-DSS for sepsis, published a pair of papers on their tool, InSight. These were retrospective analyses of patients in the MIMIC-II[61] and MIMIC-III[62] databases. In the MIMIC-II study, the focus was on hospital-acquired sepsis, so the company screened for patients that do not present with SIRS at the time of admission or within the first four hours of stay, and who had documented measurements for SBP, pulse pressure, heart rate, temperature, respiration rate, WBC, pH, Blood oxygen saturation and age (a total of 1,394 patients). Of that set, patients whose record contained an ICD9 code (995.9) for sepsis and met the SIRS criteria for a 5-hour period of time were used for the trial, a total of 159 patients. In that case, InSight was able to predict the possibility of a patient being given an ICD code and presenting with SIRS criteria 3 hours before the 'zero hour' (point of first diagnosis), to an accuracy of .92, higher than a PCT test (which had an accuracy of 0.85).[61] While promising, the sample size of 159 is somewhat inconclusive, which might suggest why the second study, based on MIMIC-III was used.

The MIMIC-III study analysed a set of 1,840 septic ICU stays, against 17,214 nonseptic patients. Interestingly, this assessment also tested the system "in the presence of data sparsity" – that is, when randomly selected pieces of information were deleted from the EHR data. Insight was also tested against a series of diagnostic criteria, such as SIRS, SOFA, MEWS (Modified Early Warning Score) and SAPS II (Simplified Acute Physiology Score II). For markers, InSight used systolic blood pressure, pulse pressure,

heart rate, respiration rate, temperature, Glasgow Coma score, and peripheral capillary oxygen saturation, arguing that "all of these features are nearly universally available at the bedside and do not rely on laboratory tests." The study found a higher performance than SAPS II and SOFA, and comparable performance with other machine learning algorithms, without requiring biomarkers.

While this information is promising, it is important to note that it is retrospective, and not directly used in a clinical trial. Dascena was later used in a prospective randomized clinical trial, in a paper published by Shimabukuro et al[53]. The intent was to measure if the use of Dascena's machine learning algorithm would reduce length of stay and mortality rate in the hospital. Similar to the MIMIC-III study, the algorithm was compared to SIRS, SOFA, qSOFA, and MEWS systems. In this case, however, in addition to the non-laboratory variables (blood pressure, respiratory rate, etc.) used in the preview test, labs such as pH, WBC count, and glucose were included. The system would then generate a risk score for severe sepsis between 1 and 100, and if the patient exceeded a score of 80, the charge nurse was notified. The sample size was again small at 142 patients, but there was a sizeable 20.6% reduction in length of stay, and had a 12.4% drop in mortality as well. Patients in the experimental (ML-aided) group also received antibiotic treatment an average of 2.76 hours earlier.[53]

Taylor et al (2016) presented a machine learning algorithm that used 20 variables, from oxygen saturation to blood urea nitrate (BUN), to make its predictions. It similarly found higher accuracy than the SIRS definition.[63] Goodman et al (2016) looked specifically at using machine learning to make timely predictions of extended-spectrum

β-lactamase (ESBL)-producing bacteria.[64] Since ESBL bacteria can hydrolyze most broad-spectrum β-lactam antibiotics[65], many antibiotic regimens have limited activity against ESBL producers[66], and awareness can help clinicians use the right antibiotics. Interestingly, the study used a form of machine learning called 'recursive partitioning' that produces a more user-friendly, interpretable decision tree, so that clinicians can read its logic more easily.[64] The study used five predictors: "history of ESBL colonization/infection, chronic indwelling vascular hardware, age ≥43 years, recent hospitalization in an ESBL high-burden region, and ≥6 days of antibiotic exposure in the prior 6 months." The finding had positive and negative predictive values above 90% using those predictors. [64]

A few studies have focused on improving algorithm design or streamlining workflow. Ghosh et al. used Hidden Markov Models (HMM) to create an analysis, attmepting to create time-to-event prediction models, that see the path to sepsis as a timed series of patterns.[67] The study only used MAP, heart rate and respiratory rate, and used that to predict, given a patient at a certain state, how likely it was that the patient was going to move towards a condition of septic shock. The study argues that the use of timed relational metadata about the three variables has an improved effect on prediction.[67] Hu et al. focused on proposing a more automated model for manual chart review, using machine learning to 'read' outcomes and look for post-operative complications such as sepsis.[68] The paper explored a few machine learning styles, suggesting ideas for future research.

Horng et al (2017) focused on using free text (such as clinician's notes in an EHR) to augment the machine learning algorithm. This retrospective study processed the free text and would then add information such as vital signs and demographic information to the algorithm. They found that adding this free text information improved discriminatory ability, as the area under the curve (AUC) increased from 0.67 to 0.86.[69]

Taneja et al (2017)[54] presented a study that granularized phases of sepsis, and looked to find the presence of biomarkers for specific stages. They combined the study of 16 non-traditional biomarkers with 15 EHR variables, to find patterns. The study argues that a "one size fits all" system for every potential sepsis patient is an incomplete picture – providing a range of biomarkers and vital sign information would allow an algorithm to more accurately pinpoint the stage of sepsis and suggest personalized treatment.[54]

Kam et al (2017) is one of the first papers to introduce the concept of deep learning into sepsis prediction. The paper presented a response to InSight, using deep feedforward networks to create an algorithm and compare it to InSight. The study used the same MIMIC-II database as one of InSight's studies, and leveraged the more complex 'long short-term memory' ability to create an algorithm that can learn sequential patterns and make conclusions from them (this is further explained below in 'Deep Learning').[70]

Cockrell et al. (2017)[71] and Peterson et al. (2018)[72] presented two related papers that paired deep learning with simulation. Cockrell et al. proposed a simplified sepsis computer simulation, which could simulate up to 7 million sepsis 'patients' by creating a virtual immune system response. Peterson et al. took that simulation system and applied a deep learning algorithm to it, and enabled reinforcement learning – that is, it allowed the

algorithm to take 'actions' on the virtual patients, and learn how effective its actions were to reduce symptoms. This led to an algorithm that would analyze the patterns of specific cytokines in patients, and would suggest a multi-cytokine mediation therapy, personalized to the patient.[72] While this needs to be validated with studies on live organisms, the study shows great promise, and is a far cry from the "one size fits all" nature of diagnostic protocols like SIRS and SOFA.

Raghu et al (2017) applied deep reinforcement learning to a retrospective analysis of the MIMIC-III database. The study then used the reinforcement "action" system to allow the algorithm to make various protocols for when to administer IV fluids and vasopressors, specifically. It would then look at the outcomes of the various patients, and adjust its protocols to maximize the number of successful outcomes. The study found that doctors that followed the algorithm's timing protocols for administering vasopressors and IV fluids ended up with the least mortalities.[73]

A GROWING TREND OF ROBUSTNESS

The papers here show a gradual increase in complexity with regards to the application of machine learning in sepsis. At first, analysis resembles statistics more than complicated machine learning, and is used more to make conclusions about clinician behaviour. As the algorithms become more complex, a trend of retrospective analysis finding predictions of equal or higher accuracy than SIRS or SOFA emerges. Later research focuses on approaching sepsis at new angles – simulating and proposing new ways of looking at stages of the disease, and tailoring personalized treatments based on a network of biomarkers. While the applications of financial and logistical feasibility need

to be taken into account, this represents a new, exciting opportunity in sepsis diagnosis and treatment – instead of the confusing, unspecific, time-intensive nature of diagnostic tools such as SIRS and SOFA, machine learning could create highly specific treatments that can look at larger volumes of data than a clinician could, and suggest specific treatments designed for specific stages of sepsis.

Of particular interest is the theorized multi-cytokine-based prediction system proposed by Peterson et al.[72] Sepsis and SIRS have both been criticized for issues with specificity, and clinician diagnosis from visually identifiable symptoms combined with a handful of commonly available biomarkers is likely always going to present an incomplete picture compared to a granular analysis of cytokine activity in the body. It will be interesting to see if the future analysis of this proposal leads to a new mode of thinking regarding sepsis diagnosis.

MACHINE LEARNING FUNDAMENTALS

A comprehensive analysis of the multiple types of machine learning is beyond the scope of this paper, and it is important to note that there are dozens of types of machine learning algorithms that one can apply to a problem such as sepsis (LASSO, naïve Bayes, random forest) that are used in machine learning approaches to sepsis detection. However, It is useful for clinicians to have an understanding of the basics of machine learning, to understand what is and is not being analyzed.

Machine learning represents a philosophical shift in the tackling of a problem. Instead of attempting to define every possible angle to a problem and code solutions for

each possibility, a machine learning program is designed to be given a set of data, and to find patterns through statistical analysis. The advantage here is that problems that cannot be simplified into a simple rule set can be tackled with machine learning.

A classic example of this is the detection of handwritten numbers[74]. Standardized fonts will always show a '9' with the same width, the same curve or line representing the lower half of the '9', but handwritten versions can vary dramatically, and programming a ruleset to cover every possible combination would be very complex. Instead, if a machine learning program is given a database of handwritten numbers, along with a 'correct answer' for which number the drawing represents, it is possible to train the program to predict the numbers to a high degree of accuracy. The program is told only that there are 10 values (0-9) and that the image it is presented with is one of those numbers. The computer makes a 'guess', then looks at the right answer (provided with the image). If it is incorrect, it modifies parts of its algorithm and tries again with the next number, until it is able to recognize numbers with a high accuracy[75].

This is, of course, an oversimplification, as there is a deep level of complexity surrounding many parts of machine learning. The actual structure of the algorithm, how it decides to make an actual guess, and how it corrects for a 'wrong' guess, are complicated topics with many avenues to explore. However, there are a few basic similarities worth delineating for a clinician to understand.

Many machine learning algorithms use a node-based system (referred to as 'neurons'), where a decision in one neuron activates the next neuron, in a manner inspired by the neuron excitation system used in the human brain[76]. While the ultimate

design is not entirely the same as the human neuronal system, the basic structure is similar. To continue with the example of handwritten digit recognition, imagine that there are only 10 possible right answers (0-9), and the algorithm is designed to conclude which of the 10 answers it is "seeing" in the image of a black '3'. In order to do so, it constructs a series of columns, each with a number of neurons. The final column consists of 10 neurons, the answers 0-9. It then looks at the information provided by the image (for instance, the black/white value of pixel 1, pixel 2, pixel 3, etc). Pixels are then grouped into neurons, and the neurons that contain black color will then 'fire' and turn on other neurons, similar to an organic neuron system. Eventually, the neurons all must settle on one of 10 possible answers, 0-9, by looking at which groups of neurons are being turned on.

However, not every piece of information will be equally relevant – for instance, the presence of hyperlactatemia is an more relavant indicator of sepsis than age or body weight, so that information must be seen as more significantly by the program. In the case of recognizing numbers from an image, perhaps corners of an image are less likely to be relevant for understanding the image than the middle area. As such, the computer must learn to 'weight' some information over others. This weighing of certain neuronss becomes key to the machine's "learning". Initially, all neurons might be weighed randomly, or equally. As the program gets more things wrong or right, it adjusts the values of these weights, and with enough repeated cycles, it finds that certain weights are more likely to produce a right answer than others. This works in a manner similar to the psychological principal of positive and negative reinforcement – the more a weight

supports correct answers, the larger it becomes. In the end, through hundreds to thousands of guesses and tiny corrections, the weights are tuned to reliably 'read' what the machine is being given.

LIMITATIONS TO MACHINE LEARNING

Due to the nature of this style of learning, there are a few strong caveats. For instance, machine learning systems are very 'data-hungry' – that is, they require large sets of data before they can start doing accurate predictions. For instance, the MNIST database of handwritten digits contains 60,000 'training' samples (for the machine learning algorithm to train on) and 10,000 'test' samples for the algorithm to attempt prediction with[77]. The amount of examples necessary to properly train a specific algorithm differs for the various types of machine learning algorithms (Markov Models, Deep Learning, Logistic Regression, etc.). The high volume of examples, however, is in many ways a product of the statistical roots of machine learning, as large volumes of data are required to accurately assess trends in the data. As changes to the weights of algorithms are made incrementally, large volumes of examples are needed in order for the program to make predictions with a high degree of accuracy.

Because of this high volume requirement, getting a large enough volume of data was previously a difficult task. However, the growing volume of electronic health records (EHR) has allowed for the automated collection of relevant data, which has allowed the regular use of machine learning for data analysis to be a viable, potentially powerful tool.[78]

Unfortunately, a second major roadblock impedes progress, even with an EHR – the problem of 'dirty' data. As the program does not necessarily know what a specific value means, it is largely unable to discern if a certain value is 'correct'. In machine learning, the learning is only as good as the data presented to it – if the data collected is missing critical information, is mislabeled, or is riddled with incorrect information, the machine will make incorrect assumptions due to having this faulty data. For data scientists, cleaning data is a large, resource-intensive part of their job – enough so that it is considered the major roadblock to progress when working on a system[79]. 'Clean' data refers to data that has been checked for missing data, which can be a very time-intensive process.

One of the issues that arises from the clinical data collected is that the quality and comprehensiveness of clinical data is at a lower level than research data.[80] This is partly due to a difference in priorities when collecting data, as research-based data gathering will have a different set of priorities than clinical data gathering.[81] Another issue is that clinical data will often have handwritten notes describing a physician's diagnosis, and those notes are generally unstructured[69] compared to standardized codes, and may contain shorthand abbreviations that are not standardized. If an important piece of information is contained in these notes, a machine learning algorithm would have to learn to take that into account. Some research has been done into natural language processing (NLP) for this purpose,[56] and some studies have used this improve Machine Learning diagnosis for sepsis.[69]

ADVANCED MACHINE LEARNING ALGORITHMS – DEEP LEARNING, HIDDEN MARKOV MODELS

In recent years, the challenges that machine learning have been able to address have gotten more and more complicated. For instance, image recognition algorithms have been programmed that can recognize skin based melanoma cancers with a dermatologist's degree of accuracy, using only a phone's camera[82]. Advancements in this vein have come about because of an increase in computing power, the commercialization of computational tools such as IBM Watson and because of further research into advanced forms of machine learning, such as Deep Learning.

HIDDEN MARKOV MODELS

Advanced systems of machine learning are designed around different paradigms, and these differences allow for predictions in different veins. For example, one approach is the Hidden Markov Model, a variant of the Markov Model (or Markov Chain). Similar to traditional methods, Markov Models look at a limited series of states (for instance, 'healthy,' 'sepsis,' 'severe sepsis', 'septic shock,' 'deceased'), and calculates the probability in of moving to a new state depending only on the current state (a rule known as the "Markov Property").[67] The intent is to model the relationship between states as a function of probability. A common example, described in Ramage 2007[83], is the modelling of weather patterns. If the current state of weather in Los Angeles is "sunny", what is the probability that it will transition within a few days to "cloudy," or "raining?" If every transition is calculated (for instance, by recording weather states for 100 days

and noting transitions), a set of data can be generated that models the chance of a certain weather state, based on the current stated.

$$A = \begin{array}{c|cccc} & s_0 & s_{sun} & s_{cloud} & s_{rain} \\ s_0 & 0 & .33 & .33 & .33 \\ s_{sun} & 0 & .8 & .1 & .1 \\ s_{cloud} & 0 & .2 & .6 & .2 \\ s_{rain} & 0 & .1 & .2 & .7 \end{array}$$

Figure 2. Sample probability table of transitions between weather states. In this set of data, if the current state is 'sun', there is an 80% chance of the next day's weather being the same state of 'sun', and a 10% chance of changing to either 'cloud

A normal Markov Model requires that the states are observable – that is, one must be able to go outside and see the weather for 100 days, in order to create the table. This presents a difficulty, therefore, if one wished to calculate this same probability, but for weather in the past when no weather data was taken. Similarly, in the case of sepsis, the condition of blood infection is similarly hidden from the clinician –testing for the presence of pathogens in the blood of a septic patient yields a positive result only 30% of the time[17] – so the clinician must assess the likelihood of the patient having sepsis based on other factors such as heart rate or respiratory rate or hyperlactatemia. In cases like this, the Hidden Markov Model is used.

In the Hidden Markov Model, one cannot observe the system they wish to model, and as such must model an ancillary, related system. In the analogy of weather, imagine if the modeler was attempting to model the weather in Los Angeles 10 years prior, and the weather data was not recorded during the required time period, but daily sales data of

umbrellas in a local open-air mall was readily accessible. Logically, sales data of umbrellas correlate to rainy days, though the conclusion is not necessarily guaranteed (perhaps on a dry day, someone needs to purchase an umbrella for a trip to another rainy city, or on a rainy day, fewer people purchase an umbrella because they bought one the rainy day before). While the information doesn't correlate directly, one could create a model of various purchasing states, and then use that to make a model predicting weather patterns.

In the case of Sepsis, then, the use of observable factors (heart rate, respiratory rate, etc.) are available for a hospitalized patient, while the 'state' of sepsis is not readily accessible. Logically, however, the factors have a correlation to sepsis, as defined by systems such as SOPA and APACHE II guidelines. As such, the rates of transition from, say, 'normal heart rate' to 'tachychardic heart rate' are observable and modelable, and can be used to predict the 'hidden' system (the probability of going from 'mildly septic' to 'severely septic' given one's current health data).

Two studies have used the Hidden Markov Model system for sepsis prediction. Peterson, et. Al used the system to predict five states – discharged, sepsis of severity level 1,2,3 (known as S1, S2, S3) and deceased – retrospectively on a collection of data from the Kaiser Permanente Northern California Data Set.[84] The model was built using five common vital signs – respiratory rate, systolic blood pressure, diastolic blood pressure, heart rate, and temperature. To give context to that data, there were three "covariate" factors used as well – age, acute physiology score (LAPS2), and chronic disease burden score (COPS2). Using the records of 20,000 patient hospitalization episodes,[84] the system

modeled a range of likely values for the five vital signs, for 'discharged', 'S1', 'S2', and so on. This data was then validated against the "Sepsis-1" SIRS criteria and qSOFA, and used to predict mortality risk, which was identified as the S3 state. Using this retrospective data, the study found the Hidden Markov Model used to be more accurately predictive of high-risk sepsis states than the sepsis-1 or qSOFA criteria.

Ghosh et al. extracted mean arterial pressure, heart rate and respiratory rate from the MIMIC-II database, and mapped their interactions to a Coupled Hidden Markov Model (CHMM). The study found that coupling these variables and creating a model of their influence on each other resulted in a statistically significant increase in prediction of septic shock compared to baseline models.[67]

Gultepe et al. used Hidden Markov Modeling to predict lactate levels, and then used that system as part of a larger machine learning algorithm.[58] Using only vital signs and WBC count, they were able to predict lactate levels to an accuracy of 0.99, and then couple that information with analysis of arterial pressure and respiratory rate to predict patient mortality in a larger machine learning algorithm. Vital signs and WBC were the 'known' markov variables, and those were used to understand the 'hidden' relationship of lactate levels.

DEEP LEARNING

As mentioned before, one of the biggest impediments to machine learning is the necessity of clean, accurate data. The data has to also be as similar as possible, to limit the potential for the machine learning algorithm to make incorrect conclusions. For instance, in the example of handwritten number recognition, a traditional machine learning algorithm

would have a harder time being as accurate if the numbers were presented at dramatically different sizes, or in different colors. The computer might conclude that an important characteristic of a '4' is the color red, while a '5' is twice the size of a '4'. Another famous example comes from image recognition – a 'standardized' set of faces would need to be of similar skin tone, with similar lighting condition, where the faces are centered and facing the camera. The more non-standardized conditions, the more difficulty the algorithm has correctly predicting accurately.

"Deep learning" attempts to address these two concerns, as the algorithms are designed to detect patterns within data that isn't necessarily standardized to a high degree, and which may contain erroneous or confusing data. Deep learning algorithms, for instance, are able to take images of cats, facing towards and away from the camera, at different times of day, and predict that the image it is seeing is that of a cat. The erroneous data of lighting conditions or size can be ignored, because of the different way that deep learning is making its calculations.

In the case of a cat, there are near universal aspects of cats - pointed ears, eyes, a nose and mouth, a tail and four legs - that when put together, constitute a cat. The 'neurons' (as mentioned above) for a Deep learning algorithm are designed to discover these smaller attributes, and then use that discovery to detect their presence in the data it is analyzing. If enough attributes of a cat are present, no matter the size or condition, the algorithm can recognize the presence of a cat in an image. The computational ability required to make this more complex analysis is far more significant, and the system is

more complex than a more traditional machine learning algorithm that does not collect a series of smaller attributes.

In theory, this is very applicable to EHR data, as EHR data is notoriously unstandardized. Many terms or shorthand phrases for the same condition differ from hospital to hospital, and some hospitals will have different protocols relating to, say, testing for sepsis. Handwritten doctor's notes about a condition could vary greatly from doctor to doctor. In theory, the advent of deep learning gives the opportunity to detect patterns in spite of these inconsistencies.

Deep learning has seen relatively few published studies on sepsis recognition, possibly due to the relative newness of deep learning as an area of study. However, a few studies have seen interesting results. A study by Kam et al. used deep learning applied to the same MIMIC-II database as the Insight study. It found higher rates of reliable diagnosis, crediting, in part, a "neural network architecture that can learn sequential patterns."[70] Other studies in deep learning focused on another area of machine learning, known as reinforcement learning.

REINFORCEMENT LEARNING

In traditional 'supervised' machine learning, a data point is given with a 'correct' answer of what that data point means, and the machine learning algorithm is designed to learn the associations. There is also 'unsupervised' learning, in which a dataset is given with no clear association with what is 'correct', with machine learning algorithm attempting to find patterns in the data. Reinforcement learning is different from those two, in that it is designed to learn from interactions – that is, a reinforcement learning

algorithm that would make ‘decisions’, and see how well the decisions maximized a ‘reward’ variable[85] (such as patient mortality). This is analogous to the psychological concept of reinforcement learning,[86] first made famous by Pavlov and his famous experiment on his dogs.

A good example of this form of machine learning is a study by Cockrell, et al., which created a simulation of virtual sepsis patients, and a study by Peterson, et al. that used reinforcement learning to conduct experimental treatment ‘actions’ on the virtual patients. The algorithm would then learn how well its treatment actions would help the virtual patients, and adjust its treatments until they maximized patient health.

Cockrell, et al. created a simplified model of 7 million virtual sepsis patients with a wide range of symptoms. The simulation, the Innate Immune Response Agent-Based Model (IIRABM), represents the dynamic of immune response, down to the temporary roles of pro- and anti-inflammatory cytokines.[71] While the simulations were “vastly simpler” than a real-world sepsis patient, the use of simplified models allows for comprehensive testing when all variables are known, and any potential learning can then be applied to the more complicated real-world cases. The intent was to present a body of potential “patients” to study with High Performance Computing (HCP) models such as deep learning, to see if it was possible to create personalized diagnoses of sepsis, to aid in personalized treatments rather than generalized antibiotic regimens.[71]

That study was then used by Peterson et al. to run a deep learning analysis. The analysis was designed to analyze if “adaptive, personalized multi-cytokine mediation can control the trajectory of sepsis and lower patient mortality.”[72] The study used the

IIRABM as a Reinforcement Learning environment. The deep learning algorithm would make an action upon the IIRABM simulation, and would record how effective the result was on the population of simulated sepsis patients. It would then continue experimenting with actions until it found effective policies to reduce mortality as much as possible.

Using this deep reinforcement learning model, the study was able to come to 0% mortality in simulated subjects. It created specific policies for patients that were personalized (patient-specific), involved mediating multiple cytokines simultaneously in coordination, and were adaptive to the virtual patient's progress over time.[72] While this is clearly a virtual system, and thus the results are not immediately generalizable to real-life patients, the model is instructive – it shows that deep learning can analyze large volumes of data to present a policy, which can then be tested more formally in clinical trials.

Another study by Raghu et al. analyzed the MIMIC-III database with reinforcement learning.[73] The MIMIC-III database represents patients that have already been treated, so dynamic actions on the populace are not possible. However, reinforcement learning could be used to "read" the medical interventions, and to understand how effective they were in maximizing the 'reward' of a healthy outcome (in this case, lowered SOFA scores). The benefit of using reinforcement learning is that the model can "infer optimal strategies from training examples that do not represent optimal behavior."[73] In observing patterns of sepsis treatment, it found an "optimal" vasopressor and IV fluid dose, arguing that lower mortality resulted from cases where physicians applied vasopressors and IV fluids in accordance to this optimal policy.[73] More research would be required before this practice could be regularly applied in clinical settings, but

this does present a data-driven approach to answer the question of the optimal sepsis treatment policy.

CLINICIAN DISTRUST AND ALGORITHMIC TRANSPARENCY

With more advanced machine learning designs comes more computational ability, and as such, the ability to use more data to more accurately predict a condition like sepsis. However, the complexity also makes any attempt to 'read' or understand the inner workings significantly more complex. For instance, while a clinician may not be able to see the infrared detection of an pulse oximeter, and does not personally calculate the ratio of red to infrared light (which translates into a measure of blood oxygen saturation via the Beer-Lambert Law[87]), the doctor could personally calculate the ratio, and can understand that the system is reliably based on this specific input. Machine learning, on the other hand, is a form of guessing and statistical probablility assigned to diagnose life-or-death conditions such as sepsis. As such, there has been an increase in skepticism regarding machine learning's somewhat random learning process. For instance, The research group AI Now published a 2017 report recommending that healthcare agencies should not use "black box" algorithms, as their use raises 'serious due process concerns'.[88]

An example of a machine learning algorithm making a logical, but problematic conclusion, was that of the relationship between asthma and pneumonia mortality.[89] A researcher for Microsoft explained that in the 1990s, they had trained an algorithm to predict mortality probability for pneumonia patients, to separate low-risk from high-risk individuals. The intent was to place high-risk patients in higher priority for treatment.

The algorithm looked at prior data, and deduced that asthmatic patients had a lower probability of mortality from pneumonia, and were placed in the low-risk tier. However, this lower probability occurred because asthma is considered a serious risk factor for pneumonia. Asthmatic patients were quicker to bring themselves to treatment when noticing difficulty breathing, and patients presenting with a history of asthma were admitted more immediately.[89] In other words, prioritizing asthmatic patients was what made the mortality rate decrease for these patients. The algorithm would have no way of knowing that, however, and the algorithm moving that classification of patient to the low-risk tier would have removed that benefit. If an algorithm is not designed to be readable, this sort of deduction cannot be recognized by a supervisor and corrected, and could have dangerous results.

Similarly, machine learning algorithms could make decisions based on information that is unethical, such as making triage decisions based on the ethnicity or gender of a patient, rather than on the patient's condition.[90] Part of this moral dilemma is being decided by legal systems - The recent General Data Protection Regulation (GDPR) passed by the European Union creates a "right to explanation,"[91] where a user can ask for an explanation of how an algorithmic decision was made about them. In response to ethical and legal pressure, machine learning researchers have been designing algorithms to include interpretability. For instance, an algorithm could display text snippets related to information used to inform decisions, so that a supervisor could deduce what part of the data is being used to make a certain decision.[92,93].

USABILITY AND ALARM FATIGUE

While multiple researchers have used machine learning retroactively to analyze patient data and diagnose sepsis, there have been relatively few proactive clinical trials. While there is promise in conducting a retroactive trial, more research is needed into effective, repeatable clinical trials to strengthen the case of using Machine Learning-Based Diagnostic Support Systems (ML-DSS) in the hospital setting. In addition to experiential robustness, an ML-DSS should also take care to integrate well into the clinical setting, rather than adding another burden for clinicians.

One issue to consider is the potential to adding to alarm fatigue. Many electronic systems already present in an Emergency Department or Intensive Care Unit are equipped with alarms meant to alert staff about potential emergencies. The volume of alerts currently used can be immense – A study at UCSF found that in one month, more than 2.5 million alarms were raised, and that nearly half of these alarms were for a wildly inaccurate arrhythmia alert system. Of the 1.15 million arrhythmia alarms that were generated, more than 88% were false positives.[94]

Interestingly, this problem has also attracted some interest from the machine learning community - a research group is using a machine learning system to find and suppress false alarms for arrhythmias.[95] Researchers have reported that monitor alerts result in changes in patient management less than 1% of the time[96]due to the high volume of alarms that do not convey meaningful information – if an ML-DSS program is to be

effective in the treatment of sepsis, it cannot simply add to the noise affecting clinicians in the hospital setting.

Since standards for sepsis are controversial and not entirely standardized, there is also a possibility that an ML-DSS program could be making decisions based on older, outdated guidelines. Machine learning algorithms already face distrust by medical professionals due to the the "black box" nature of the algorithm's decision process[97] – if a standard for sepsis diagnosis or billing codes were to change in five years, and the machine is not updated for the new standards, this would only add to alarm fatigue, and it is likely that staff will be similarly dismissive of any new system, no matter how theoretically effective it might be.

FUTURE IMPLEMENTATIONS

Perhaps, with proper user experience design, an ML-DSS could be used to save time, as opposed to handing a clinician more responsibilities. For instance, upon sensing that a patient's vital signs are leading towards possible sepsis, perhaps an ML-DSS could automatically order blood work for biomarkers such as PCT and CRP. If conditions continue to worsen, the system could automatically order tests for a more specific diagnosis, such as markers for the presence of certain bacteria. If a clinician then comes to assess the patient, a clear timetable of suspicions, of laboratory results and suspected pathogens could be presented to the clinician, simplifying the amount of steps needed to reach a diagnosis. The system could be designed such that alerts only present to clinicians once standard steps have been taken, reducing the workload for the clinician.

Machine learning algorithms have the ability to continuously process data at a level clinicians are unable to. As Taneja et al. found, machine learning algorithms can learn to prioritize certain biomarkers at certain stages, and predict the severity of sepsis due to this. Monitoring a comprehensive suite of biomarkers, especially if the ML-DSS could also automatically order blood tests, could provide a more clear picture, rather than simply alerting a clinician to the possibility of a diagnosis. It is possible that, if an ML-DSS is designed with interpretability in mind, that the interface could display the reasoning for the actions it suggests taking, providing a sort of upgraded, commented lab result that a clinician could use to make decisions.

Similarly, ML-DSS programs could be used to reduce patient rehospitalization. Rehospitalizations for patients with sepsis are mostly due to infection or another septic episode. After an index sepsis incident, nearly 40% of patients are transferred to skilled nursing facilities.[33] Perhaps an ML-DSS program could be designed to communicate with facility systems to provide similar diagnostic support, and monitor for signs of infection, prompting recommendations for antibiotic treatment to be administered at the skilled nursing facility.

CONCLUSION

Sepsis is a complex, deadly disease, known since the time of Hippocrates.[98] In the present day, it is one of the leading causes of mortality in the hospital setting. It is also not fully understood – pathologies of sepsis differ depending on the age of the patient and prior comorbidities, making diagnosis difficult. Bloodwork can't paint a complete picture, as there's no single biomarker that clearly delineates sepsis, and the most prevalently tested

biomarkers are chosen for laboratory availability rather than specificity. The very definition of sepsis in the medical community is controversial, subject to a major change in 2016, which is not universally accepted. Patient factors are also highly influential with respect to outcome and not always easily mineable from the chart. Patients that contract sepsis live with a severe symptoms, prone to re-hospitalization (often due to infection) and with a five times higher chance of mortality five years after the index event. Given the high cost of life and resources, there is a real need to address sepsis and minimize the loss of life and limb.

Traditional computerized 'alerts' that monitor vital signs and notify clinicians of possible sepsis have been around for some time, but they are limited to only reporting a patient already in sepsis. Academic research has been investigating using machine learning to *predict* sepsis instead, and to increase the accuracy and timeliness of both sepsis diagnosis and treatment. The field has seen multiple promising studies, with algorithms that are reported to be equally as sensitive and specific as current diagnostic protocols such as SIRS and qSOFA. Some algorithms even propose new methods of detection and treatment, looking at a more granular level at the levels of individual cytokines on a continuous level, to guage how far along the condition of sepsis a patient is. These proposed algorithms have lots of promise for prediction, diagnosis and treatment, but more research needs to be conducted before these studies are to be seen in Emergency Departments around the world. Most of the studies conducted so far are retrospective studies done on EHR databases, and few prospective studies have been performed. However, this is not far off – one company, Dascena, has tested Machine

Learning-Based Diagnostic Support System (ML-DSS) called InSight in a live clinical trial. To the author's knowledge, InSight is the only commercially available ML-DSS tool currently available, but more are likely to follow.

A few complications need to be considered for an ML-DSS to be truly effective. An ML-DSS should be transparent and intuitive enough that a clinician can understand what the decisions are being based off of, so that the clinician can trust the software, and can check for erroneous assumptions that the algorithm is making. In addition, for true acceptance in a clinical setting, an ML-DSS should reduce, rather than add to, the alert fatigue that is currently present in hospitals. Adding another alert to the cacophony being given to a clinician will likely not help in the timely diagnosis and treatment of sepsis. However, if proper thought is put into the design of an ML-DSS for a clinical setting, and financial and logistical considerations are further developed, machine learning can have a powerful, useful, and ultimately life-saving role in the diagnosis and treatment of sepsis.

REFERENCES

1. Tromp, M. *et al.* The role of nurses in the recognition and treatment of patients with sepsis in the emergency department: a prospective before-and-after intervention study. *International Journal of Nursing Studies* 47, 1464–1473 (2010).
2. Angus, D. C. *et al.* Epidemiology of severe sepsis in the United States: Analysis of incidence, outcome, and associated costs of care. *Critical Care Medicine* 29, 1303 (2001).
3. Angus, D. C. & van der Poll, T. Severe Sepsis and Septic Shock. *http://dx.doi.org/10.1056/NEJMra1208623* (2013). doi:10.1056/NEJMra1208623
4. Angus, D. C., Carlet, J. & 2002 Brussels Roundtable Participants. Surviving intensive care: a report from the 2002 Brussels Roundtable. *Intensive Care Medicine* 29, 368–377 (2003).
5. Simpson, S. Q. SIRS in the Time of Sepsis-3. *Chest* 153, 34–38 (2018).
6. Thomas-Rueddel, D. O. *et al.* Hyperlactatemia is an independent predictor of mortality and denotes distinct subtypes of severe sepsis and septic shock. *Journal of Critical Care* 30, 439.e1-439.e6 (2015).
7. Singer, M. *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315, 801–810 (2016).
8. Shankar-Hari, M. & Rubenfeld, G. D. Understanding Long-Term Outcomes Following Sepsis: Implications and Challenges. *Current Infectious Disease Reports* 18, (2016).

9. Weycker, D., Akhras, K. S., Edelsberg, J., Angus, D. C. & Oster, G. Long-term mortality and medical care charges in patients with severe sepsis. *Critical Care Medicine* 31, 2316–2323 (2003).

10. Bloos, F. & Reinhart, K. Rapid diagnosis of sepsis. *Virulence* 5, 154–160 (2014).

11. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016 | SpringerLink. Available at: https://link-springer-com.ezproxy.bu.edu/article/10.1007/s00134-017-4683-6. (Accessed: 19th April 2018)

12. Cao, Z. & Robinson, R. A. S. The role of proteomics in understanding biological mechanisms of sepsis. *PROTEOMICS – Clinical Applications* 8, 35–52

13. Cohen, J. The immunopathogenesis of sepsis. *Nature* (2002). doi:10.1038/nature01326

14. Schouten, M., Wiersinga, W. J., Levi, M. & Poll, T. van der. Inflammation, endothelium, and coagulation in sepsis. *Journal of Leukocyte Biology* 83, 536–545

15. Remick, D. G. Pathophysiology of Sepsis. *The American Journal of Pathology* 170, 1435–1444 (2007).

16. Bellos, I. *et al.* Soluble TREM-1 as a predictive factor of neonatal sepsis: a meta-analysis. *Inflammation Research* 1–8 (2018). doi:10.1007/s00011-018-1149-4

17. Calandra, T., Cohen, J. & International Sepsis Forum Definition of Infection in the ICU Consensus Conference. The international sepsis forum consensus conference on definitions of infection in the intensive care unit. *Critical Care Medicine* 33, 1538–1548 (2005).

18. Rothman, M. *et al.* Sepsis as 2 problems: Identifying sepsis at admission and predicting onset in the hospital using an electronic medical record–based acuity score. *Journal of Critical Care* 38, 237–244 (2017).

19. Marik, P. E. & Taeb, A. M. SIRS, qSOFA and new sepsis definition. *Journal of Thoracic Disease* 9, 943–945 (2017).

20. Tedesco, E. R., Whiteman, K., Heuston, M., Swanson-Biearman, B. & Stephens, K. Interprofessional Collaboration to Improve Sepsis Care and Survival Within a Tertiary Care Emergency Department. *Journal of Emergency Nursing* 43, 532–538 (2017).

21. Bruce, H. R., Maiden, J., Fedullo, P. F. & Kim, S. C. Impact of Nurse-Initiated ED Sepsis Protocol on Compliance With Sepsis Bundles, Time to Initial Antibiotic Administration, and In-Hospital Mortality. *Journal of Emergency Nursing* 41, 130–137 (2015).

22. Bone, R. C. *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 101, 1644–1655 (1992).

23. Wacker, C., Prkno, A., Brunkhorst, F. M. & Schlattmann, P. Procalcitonin as a diagnostic marker for sepsis: a systematic review and meta-analysis. *The Lancet. Infectious Diseases* 13, 426–435 (2013).

24. Nguyen, S. Q. *et al.* Automated electronic medical record sepsis detection in the emergency department. *PeerJ* 2, (2014).

25. Pierrakos, C. & Vincent, J.-L. Sepsis biomarkers: a review. *Critical Care* 14, R15 (2010).

26. Do We Need Procalcitonin for Sepsis? - AACC.org. Available at: https://www.aacc.org/publications/cln/articles/2015/february/procalcitonin-for-sepsis. (Accessed: 23rd June 2018)

27. Mimoz, O. *et al.* Procalcitonin and C-reactive protein during the early posttraumatic systemic inflammatory response syndrome. *Intensive Care Medicine* 24, 185–188 (1998).

28. Meisner, M., Tschaikowsky, K., Hutzler, A., Schick, C. & Schüttler, J. Postoperative plasma concentrations of procalcitonin after different types of surgery. *Intensive Care Medicine* 24, 680–684 (1998).

29. Pepys, M. B. & Hirschfield, G. M. C-reactive protein: a critical update. *Journal of Clinical Investigation* 111, 1805–1812 (2003).

30. Lobo, S. M. A. *et al.* C-reactive protein levels correlate with mortality and organ failure in critically ill patients. *Chest* 123, 2043–2049 (2003).

31. Liaudat, S., Dayer, E., Praz, G., Bille, J. & Troillet, N. Usefulness of procalcitonin serum level for the diagnosis of bacteremia. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 20, 524–527 (2001).

32. Torio, C. M. & Moore, B. J. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2013: Statistical Brief #204. in *Healthcare Cost and*

Utilization Project (HCUP) Statistical Briefs (Agency for Healthcare Research and Quality (US), 2006).

33. Chang, D. W., Tseng, C.-H. & Shapiro, M. F. Rehospitalizations Following Sepsis: Common and Costly. *Critical care medicine* 43, 2085–2093 (2015).

34. Dietz, B. W., Jones, T. K., Small, D. S., Gaieski, D. F. & Mikkelsen, M. E. The Relationship Between Index Hospitalizations, Sepsis, and Death or Transition to Hospice Care During 30-Day Hospital Readmissions. *Medical Care* 55, 362 (2017).

35. Sun, A. *et al.* Association Between Index Hospitalization and Hospital Readmission in Sepsis Survivors. *Critical Care Medicine* 44, 478–487 (2016).

36. Braun, L., Riedel, A. A. & Cooper, L. M. Severe sepsis in managed care: analysis of incidence, one-year mortality, and associated costs of care. *Journal of managed care pharmacy: JMCP* 10, 521–530 (2004).

37. Goodwin, A. J., Rice, D. A., Simpson, K. N. & Ford, D. W. Frequency, Cost and Risk Factors of Readmissions among Severe Sepsis Survivors. *Critical care medicine* 43, 738–746 (2015).

38. Churpek, M. M., Zadravecz, F. J., Winslow, C., Howell, M. D. & Edelson, D. P. Incidence and Prognostic Value of the Systemic Inflammatory Response Syndrome and Organ Dysfunctions in Ward Patients. *American Journal of Respiratory and Critical Care Medicine* 192, 958–964 (2015).

39. Kaukonen, K.-M., Bailey, M., Pilcher, D., Cooper, D. J. & Bellomo, R. Systemic inflammatory response syndrome criteria in defining severe sepsis. *The New England Journal of Medicine* 372, 1629–1638 (2015).

40. Bermejo-Martin, J. F., Martín-Fernandez, M. & Almansa, R. Pre-sepsis: A necessary concept to complete the SEPSIS-3 picture? *Journal of Critical Care* 44, 148 (2018).

41. Chamberlain, N. Sepsis and Septic Shock. Available at https://www.atsu.edu/faculty/chamberlain/website/lectures/lecture/sepsis2007.htm

42. Teasdale, G. *et al.* The Glasgow Coma Scale at 40 years: standing the test of time. *The Lancet. Neurology* 13, 844–854 (2014).

43. qSOFA :: quick Sepsis Related Organ Failure Assessment. Available at: http://www.qsofa.org/. (Accessed: 30th June 2018)

44. Williams, J. M. *et al.* Systemic Inflammatory Response Syndrome, Quick Sequential Organ Function Assessment, and Organ Dysfunction: Insights From a Prospective Database of ED Patients With Infection. *Chest* 151, 586–596 (2017).

45. Tan, V. S. R., Phua, J. & Lim, T. K. In sepsis, beyond adherence, timeliness matters. *Journal of Thoracic Disease* 9, 2808–2811 (2017).

46. Alhashem, F., Tiren-Verbeet, N. L., Alp, E. & Doganay, M. Treatment of sepsis: What is the antibiotic choice in bacteremia due to carbapenem resistant Enterobacteriaceae? *World Journal of Clinical Cases* 5, 324–332 (2017).

47. Kalra, O. P. & Raizada, A. Approach to a Patient with Urosepsis. *Journal of Global Infectious Diseases* 1, 57–63 (2009).

48. Seymour, C. W. *et al.* Time to Treatment and Mortality during Mandated Emergency Care for Sepsis. *The New England Journal of Medicine* 376, 2235–2244 (2017).

49. Sawyer, A. M. *et al.* Implementation of a real-time computerized sepsis alert in nonintensive care unit patients*: *Critical Care Medicine* 39, 469–473 (2011).

50. Luo, W. *et al.* Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *Journal of Medical Internet Research* 18, (2016).

51. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* 349, 255–260 (2015).

52. Nemati, S. *et al.* An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Critical Care Medicine* 46, 547–553 (2018).

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

54. Taneja, I. *et al.* Combining Biomarkers with EMR Data to Identify Patients in Different Phases of Sepsis. *Scientific Reports* 7, (2017).

55. Tang, C. H. H. *et al.* Non-invasive classification of severe sepsis and systemic inflammatory response syndrome using a nonlinear support vector machine: a preliminary study. *Physiological Measurement* 31, 775 (2010).

56. Ohno-Machado, L. Realizing the full potential of electronic health records: the role of natural language processing. *Journal of the American Medical Informatics Association : JAMIA* 18, 539 (2011).

57. Convertino, V. A. *et al.* Use of Advanced Machine-Learning Techniques for Noninvasive Monitoring of Hemorrhage. *Journal of Trauma and Acute Care Surgery* 71, S25 (2011).

58. Gultepe, E. *et al.* From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *Journal of the American Medical Informatics Association* 21, 315–325 (2014).

59. Mani, S. *et al.* Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *Journal of the American Medical Informatics Association* 21, 326–336 (2014).

60. Tsoukalas, A., Albertson, T. & Tagkopoulos, I. From Data to Optimal Decision Making: A Data-Driven, Probabilistic Machine Learning Approach to Decision Support for Patients With Sepsis. *JMIR Medical Informatics* 3, (2015).

61. Calvert, J. S. *et al.* A computational approach to early sepsis detection. *Computers in Biology and Medicine* 74, 69–73 (2016).

62. Desautels, T. *et al.* Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Medical Informatics* 4, e28 (2016).

63. Taylor, R. A. *et al.* Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven, Machine Learning Approach. *Academic Emergency Medicine* 23, 269–278 (2016).

64. Goodman, K. E. *et al.* A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum β-Lactamase-Producing Organism.

Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America 63, 896–903 (2016).

65. Paterson, D. L. & Bonomo, R. A. Extended-Spectrum β-Lactamases: a Clinical Update. *Clinical Microbiology Reviews* 18, 657–686 (2005).

66. Tamma, P. D. *et al.* Carbapenem Therapy Is Associated With Improved Survival Compared With Piperacillin-Tazobactam for Patients With Extended-Spectrum β-Lactamase Bacteremia. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 60, 1319–1325 (2015).

67. Ghosh, S., Li, J., Cao, L. & Ramamohanarao, K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *Journal of Biomedical Informatics* 66, 19–31 (2017).

68. Hu, Z. *et al.* Accelerating Chart Review Using Automated Methods on Electronic Health Record Data for Postoperative Complications. *AMIA Annual Symposium Proceedings* 2016, 1822–1831 (2017).

69. Horng, S. *et al.* Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS ONE* 12, (2017).

70. Kam, H. J. & Kim, H. Y. Learning representations for the early detection of sepsis with deep neural networks. *Computers in Biology and Medicine* 89, 248–255 (2017).

71. Cockrell, C. & An, G. Sepsis Reconsidered: Identifying novel metrics for behavioral landscape characterization with a high-performance computing implementation of an agent-based model. *Journal of theoretical biology* 430, 157–168 (2017).

72. Petersen, B. K. *et al.* Precision medicine as a control problem: Using simulation and deep reinforcement learning to discover adaptive, personalized multi-cytokine therapy for sepsis. *arXiv:1802.10440 [cs, q-bio]* (2018).

73. Raghu, A. *et al.* Deep Reinforcement Learning for Sepsis Treatment. *arXiv:1711.09602 [cs]* (2017).

74. But what *is* a Neural Network? | Chapter 1, deep learning - YouTube. Available at: https://www.youtube.com/watch?v=aircAruvnKk&t=1s. (Accessed: 29th May 2018)

75. Nielsen, M. A. Using neural nets to recognize handwritten digits. (2015). Available at http://neuralnetworksanddeeplearning.com/chap1.html

76. Draelos, T. J. *et al.* Neurogenesis Deep Learning. *arXiv:1612.03770 [cs, stat]* 526–533 (2017). doi:10.1109/IJCNN.2017.7965898

77. MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges. Available at: http://yann.lecun.com/exdb/mnist/. (Accessed: 12th June 2018)

78. Adkins, D. E. Machine Learning and Electronic Health Records: A Paradigm Shift. *American Journal of Psychiatry* 174, 93–94 (2017).

79. The State of ML and Data Science 2017. *Kaggle* Available at: https://www.kaggle.com/surveys/2017. (Accessed: 8th June 2018)

80. Weiner, M. G. Toward Reuse of Clinical Data for Research and Quality Improvement: The End of the Beginning? *Annals of Internal Medicine* 151, 359 (2009).

81. Weiskopf, N. G. & Weng, C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association : JAMIA* 20, 144–151 (2013).

82. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 115–118 (2017).

83. Ramage, D. Hidden Markov Models Fundamentals. 13

84. Petersen, B. K. *et al.* Modeling sepsis progression using hidden Markov models. (2018).

85. Press, T. M. Reinforcement Learning. *The MIT Press* Available at: https://mitpress.mit.edu/books/reinforcement-learning. (Accessed: 1st July 2018)

86. Glimcher, P. W. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences* 108, 15647–15654 (2011).

87. Anaesthesia UK : Principles of pulse oximetry. Available at: http://www.frca.co.uk/article.aspx?articleid=332. (Accessed: 25th June 2018)

88. AI Now 2017 Symposium. Available at: https://ainowinstitute.org/events/2017-symposium.html. (Accessed: 2nd July 2018)

89. Caruana, R. Intelligible Machine Learning for Critical Applications Such As Health Care. in (aaas, 2017).

90. Char, D. S., Shah, N. H. & Magnus, D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *The New England journal of medicine* 378, 981–983 (2018).

91. Goodman, B. & Flaxman, S. European Union regulations on algorithmic decision-making and a ‘right to explanation’. *AI Magazine* 38, 50 (2017).

92. Lei, T., Barzilay, R. & Jaakkola, T. Rationalizing Neural Predictions. in 107–117 (Association for Computational Linguistics, 2016). doi:10.18653/v1/D16-1011

93. Lei, T. Interpretable Neural Models for Natural Language Processing. 119

94. Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0110274. (Accessed: 12th June 2018)

95. Ansari, S., Belle, A., Ghanbari, H., Salamango, M. & Najarian, K. Suppression of false arrhythmia alarms in the ICU: a machine learning approach. *Physiological Measurement* 37, 1186 (2016).

96. Graham, K. C. & Cvach, M. Monitor Alarm Fatigue: Standardizing Use of Physiological Monitors and Decreasing Nuisance Alarms. *American Journal of Critical Care* 19, 28–34 (2010).

97. Castelvecchi, D. Can we open the black box of AI? *Nature News* 538, 20 (2016).

98. Funk, D. J., Parrillo, J. E. & Kumar, A. Sepsis and septic shock: a history. *Critical Care Clinics* 25, 83–101, viii (2009).

VITA