A CAUSAL INFERENCE APPROACH TO PREDICT SEPSIS IN PATIENTS USING Q-LEARNING

A Thesis Proposal

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Master of Science
by

Noah Michael Dunn
Miami University
Oxford, Ohio

2020 (current Year)

Advisor: Dr. Dhananjai Rao
Reader: Dr. Vaskar Raychoudhury
Reader: Dr. Philippe Giabbanelli

This Thesis titled

A CAUSAL INFERENCE APPROACH TO PREDICT SEPSIS IN PATIENTS USING Q-LEARNING

by

Noah Michael Dunn

has been approved for publication by

The College of Engineering and Computing

and

The Department of Computer Science & Software Engineering

Advisor's Name

First Reader

Second Reader

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

In the age of the big data revolution, machine learning and advanced computational statistics dominate as a means of understanding the significance of large batches of data [7]. Domains outside of computer science are relying on an understanding of complex algorithms to improve the quality of life across sectors, and this in particular is impacting modern medical research. Hospitals and particularly ICUs are often crowded with more patients than they are capable of handling in a reasonable amount of time [8]. In cases of serious medical emergencies, time not spent working directly with a patient can result in more dangerous complications, and even death. As a result, many researchers are exploring the domain of computer-assisted diagnosis to gain as much information on patients as quickly as possible [9]. Doctors that are provided even a glimpse at the full scope of their patients' circumstances and conditions in the form of computer-generated predictions will be able to make more accurate judgement calls on distribution of available resources to patients most in need.

1.2 Contributions

The usage of computer-aided modeling, or even more specifically machine-learning related modeling to diagnose patients is not a new idea or field of study. Various avenues of prediction for different demographics of patients, available information, and modeling strategies are all being sought in parallel. Specifically, the research here is intended to contribute to the existing knowledge base in the following ways:

1. Build a model that makes use of Q-Learning and patient data to predict Sepsis
2. Optimize that model by use of Causal Inference techniques.

Chapter 2

Background & Related Work

2.1 Sepsis

2.1.1 Defining and Understanding Sepsis

According to the Third International Consensus Definitions for Sepsis and Sepsis Shock, commonly referred to as "Sepsis-3", *Sepsis* is defined as "life threatening organ dysfunction caused by a dysregulated host response to infection [10]."In layman's terms, Sepsis is the "body's extreme response to an infection" which often occurs when "an infection you already have – in your lungs, urinary tract, or somewhere else – triggers a chain reaction throughout your body [11]."

Sepsis has received attention for the past 30 years more so than ever as it is quickly being identified as one of the "largest causes of health loss worldwide" [12]. In 2017 alone, Sepsis totalled *48.9 million cases* worldwide and accounted for more than *11.0 million deaths across the globe* (with a 95% confidence interval) [12]. Because of this, Sepsis stands ranked in the top 10 causes of death worldwide [13]. Needless to say, now more than ever, Sepsis has been identified as a health concern worth investigating.

2.1.2 Septic Shock

An extreme form of Sepsis is known as *Septic Shock* which occurs when the body-wide infection due to sepsis "...leads to dangerously low blood pressure..." [14]. In recent years, work on understanding the causes behind Septic Shock and various potential treatments have reached an all-time high. On a microbiological level, small *Pathogens*, which consist of any organism that "[cause] disease to [their] host", initialize a series of communications among a host's cells [15][16]. The communication chain prompts the host's cells to become inflammatory or get bigger, across the entirety of the host's body. As a result, in a very short amount of time, a large portion of a person's cells suffer inflammation, and all of these cells in combination deal severe damage to the person's bodily tissue [16].

The reason that Septic Shock poses such an issue is that "The time window for interventions is short..." and any supplied treatment must be delivered soon after the infection begins to spread [16]. Septic Shock is identified as being one of the most deadly forms of Sepsis, making up an average mortality rate of *50%* [17]. At an average of *2-3%* of patients who come into the ICU identified as having Septic Shock, this amounts to a significant number of deaths caused by the condition.

2.1.3 Symptoms of Sepsis and Risk Groups

For doctors, the apparent difficulty in the detection of Sepsis the ambiguity of the signs and symptoms associated with it. All the symptoms are based on the underlying pathogen that is infecting specific parts of the affected person's body. Additionally, early genetic testing in patients has been shown to yield contradictory results [18]. To make matters additionally complicated, brain function is often "deranged" due to Sepsis impairing brain functionality in affected patients, even ones with otherwise normal brain function.

In terms of risk assessment, there are several groups of people who are more at risk of getting Sepsis. The following is a short list provided by WebMD [19]:

- People who are currently suffering from or have suffered from HIV/AIDS or Cancer
- People who take immune-system suppressors like steroids or drugs used to prevent "rejection of transplanted organs"
- Very young babies (3 months old or younger)
- The elderly (particularly those suffering any other health issues)
- Anyone who has recently suffered a major injury
- People with diabetes

Affliction	Known Symptoms + Signs
Sepsis	A fever of above 101 °F A heart rate higher than 90 BPM Breathing rate higher than 20 breaths per minute Probable or Confirmed Infection
Severe Sepsis	Any of the following and: Patches of discolored skin Decreased Urination Changes in mental ability Low platelet (blood clotting cells) count Problems breathing Abnormal heart functions Chills due to fall in body temperature Unconsciousness Extreme weakness
Septic Shock	Any of the previous and extremely low blood pressure

Table 2.1: Symptoms of the Various Stages of Sepsis [1]

If any of these kinds of patients display any of the symptoms associated with Sepsis, they will be given tests that can provide helpful information to a doctor. Among initial tests,

which determine "changes in body temperature, leukocyte count, heart rate, blood pressure, and respiration rate" are used to detect inflammation in a patient's body [19]. Following this, a doctor may use a version of the SOFA score to determine the likelihood that a patient has Sepsis given their vitals.

2.1.4 SOFA Score

The original Sequential Organ Failure Assessment (SOFA) score was developed by the European Society of Intensive Care Medicine back in 1994 [20]. In general, this score is used to determine how in danger a patient is of organ failure, due to Sepsis-related complications. The score *ranges from zero to four for each of six different organ systems.* A zero on the SOFA score represents a normal functioning organ, and a four represents a very abnormal organ [21]. Although Sepsis is not directly determined from any given case of organ failure, "...mortality rate is directly related to the degree of organ dysfunction [21]." The authors of the SOFA score make a careful note to point out here, "...SOFA score is designed not to predict outcome but to describe a sequence of complications in the critically ill.." and as such, the '...SOFA score does not compete with the existing severity indexes but complements them [21]."

2.1.5 qSOFA Score

The former SOFA score, albeit thorough in scope, does not lend itself to be performed in a quick manner. As a result, the quickSOFA score commonly known as the *qSOFA score* was developed at the Sepsis-3 conference discussed earlier [10]. The design of this score was to get a much quicker idea of whether or not a patient had Sepsis. As opposed to the default SOFA score, which has 6 different metrics of analysis, one for each organ, the qSOFA score offers just one range of scoring. The qSOFA score ranges from values of *zero to three points, where two points or more represents "...a greater risk of death or prolonged intensive care unit stay"*[3]. In qSOFA testing, a patient adds a point to their default score of zero if any of the following three criteria apply to them [3].

1. Patient has low blood pressure, defined as a Systolic Blood Pressure of $\leq$ 100 mm Hg [22]

2. Patient has a high respiratory rate, defined as $\geq$ 22 breaths/minute

3. Patient possesses an altered mentation (change in mental ability), defined as scoring $<$ 15 on the Glasgow coma scale [23]

The qSOFA Score test requires three easy to obtain measures of a patient's health, and particularly in cases outside the ICU, "...the simple qSOFA model performed similarly to more complex models like SOFA...outside the ICU [3]." The following figure shows risk relative to qSOFA score.

Component Tested	Score
Eye Response	
Eyes open spontaneously	4
Eye opening to verbal command	3
Eye opening to pain	2
No eye opening	1
Motor Response	
Obeys Command	6
Localises Pain	5
Withdraws from Pain	4
Flexion Response to Pain	3
Extension Response to Pain	2
No Motor Response	1
Verbal Response	
Oriented	5
Confused	4
Inappropriate Words	3
Incomprehensible Sound	2
No Verbal Response	1

Table 2.2: The Glasgow Coma Scale [2]

2.1.6 The Importance of Quick Treatment

After a patient is diagnosed, either by using a SOFA score, a qSOFA score, or some other metric, there is a very small window a patient has to receive treatment before their condition worsens quickly. For sepsis treatment, any "Early goal directed therapy completed within the *first six hours of sepsis recognition* significantly decreases in-hospital mortality [24]." In a study conducted by Kumar A et al., each hour of delay in treatment yielded an *average decrease in survival rate of 7.6%* [25]. This study also confirms the median time to effective treatment was six hours. In addition, the authors make note that time to effective treatment was the greatest predictor of patient outcome. Despite all of this, "only *50%* of septic shock patients received *effective antimicrobial therapy within 6 hours of documented hypotension*" (low blood pressure) [25]. A flowchart mapping the full understanding of Sepsis diagnosis and treatment is shown in *A.1*

2.1.7 Understanding Sepsis Treatment

Quick treatment is a necessity for all potential Sepsis patients, and unlike the symptoms which are varied, the treatments are more streamlined. The CDC recommends patients who have been diagnosed with Sepsis to be treated in the following manner [26]:

- The patient should receive *antibiotics*

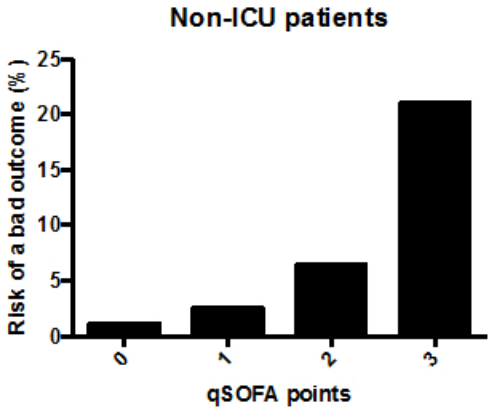


Figure 2.1: qSOFA Risk Assessment in a Non-ICU environment [3]

- Consistent blood flow should be provided to the patient's organs. This is often done through the provision of *Oxygen and Intravenous (IV) fluids to the organs.*
- The source of the infection should be treated

In a study conducted by Seymour CW et. al, the researchers add that *quicker administration of antibiotics* were "..associated with low risk-adjusted in-hospital mortality", but quicker completion of IV fluids did not follow this trend. Dr. Susan Duffy, MD, MPH, also offers a graph shown in the following figure to demonstrate the importance of early antibiotic administration in pediatric sepsis [4]. This graph also demonstrates how few patients get the treatment they need as the rate that they should.

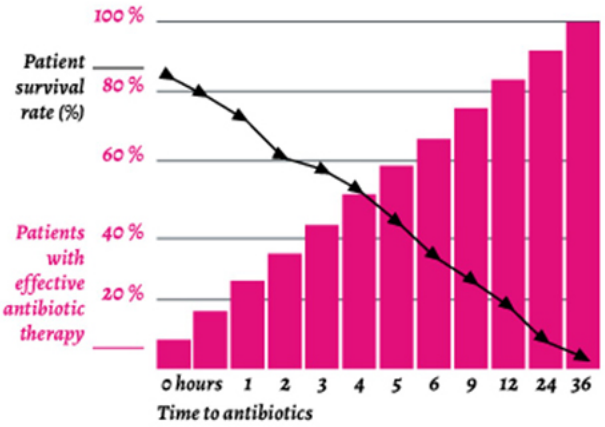


Figure 2.2: The Importance of Speed in Antibiotic Administration [4]

2.2 Q-Learning

The SOFA and qSOFA scores offer certain criteria that enable doctors to diagnose, with reasonable certainty, if a patient is suffering from Sepsis or not. However, it is possible that there are even better metrics to predict Sepsis in a patient if provided enough specific information about a patient's vitals. Such ideas are what prompted researchers like Komorowski et al., to pursue the creation of an AI clinician, as well as prompting PsyioNet's "Early Prediction of Sepsis from Clinical Data" challenge [27][28]. Both of these are discussed in detail later on in *Section 2.4*, as they are what inspired this research in the first place. In pursuit of manipulating known data to approximate outcomes, Machine Learning offers a worthy tool.

2.2.1 The Philosophy of Machine Learning

In his book, Machine Learning 2nd Edition, Marsland states that *"Machine learning ... is about making computers modify or adapt their actions ... so that these actions get more accurate, where accuracy is measured by how well the chosen actions reflect the correct ones* [29]." In the case of predicting Sepsis, the goal is to design a program that can take many patients' vitals information as data input, and output whether or not they have Sepsis with accuracy. Within the domain of Machine Learning, there are several different types of learning algorithms that can be used [29]:

- *Supervised Learning:* The algorithm is provided with a preset set of input data with known results (successes), the algorithm uses these to create a model that will give the correct answer for all the provided preset information. Also called "learning from exemplars".

- *Unsupervised Learning:* The algorithm is not provided with *labels(responses)*, but the algorithm attempts to create a model based on the similarities that the inputs have in common. Also called "density estimation".

- *Reinforcement Learning:* The algorithm is told when it comes up with the incorrect response, but not why the response is incorrect. From this point, it tries multiple approaches until it gets the response correct. Also called "learning with a critic".

- *Evolutionary Learning:* The algorithm uses a metric for "fitness", similar to biological evolution, to determine how good a given model it creates is. It adapts to the models that have higher fitness values.

2.2.2 Reinforcement Learning

In particular for medical purposes, several research groups have used reinforcement learning for data that may produce results in a "sequence of states", such as data that deals with treatments or any data involving time [30][31][32]. In a recent study by Zhang et al. a form

of reinforcement learning was used to optimize a treatment regiment for patients of drastically different characteristics [33]. Their approach, which fits linear models offers an easy interpretation (binary choices) and can be performed in most software groupings; however, it suffers when a linear model is chosen to map non-linear behavior. In studies of more generalized diagnosis, Ling et al. experimented with using entire descriptions of a patient's information with Reinforcement Learning algorithms to diagnose. [34] Their approach uses MDPs and a systematic processing of written English statements to attempt diagnosis on a patient. The research group also makes note of their use of the Deep Q-Learning Network approach, substituting a traditional value function with a approximation. Specific to the issue of diagnosing Sepsis, Komorowski et al. approached the problem using a combination of *Q-Learning and TD-learning*, expanded on further in the related works [27].

2.2.3 An Introduction to Q-Learning

In 1989, CJCH Watkins published a paper called "Learning from delayed rewards", which tackled the approach of previous reinforcement algorithms, but with a twist [35]. In his paper, Watkins proposed only rewarding the algorithm if it accomplishes the goal it was oriented towards. Later, Watkins would go on to formalize his Machine Learning proposal under the name *"Q-Learning"* alongside several other researchers. While a more generalized algorithm would look for ways to improve the result from the initial test and response of a model, Q-Learning seeks to optimize individual *"qualities"* (this is where the Q comes from) of a model through incremental experimentation [35][36][1]. Additionally, it is important to understand that Q-Learning is in *Off-Policy* algorithm. What this means is that at any given incremental step, the Q-Learning algorithm *does not* take into consideration *what previous path (policy)* was taken to achieve the desired outcome [2].

2.2.4 Markov Decision Processes

Since we are concerned exclusively with decisions made that do not take prior history into consideration, the underlying structure to any Q-Learning algorithm makes use of the mathematical construct of the *Markov Decision Process(MDP) model* [37]. Any given MDP contains the following [37]:

1. "A set of possible world states"
2. "A set of possible actions"
3. "A real valued reward function"
4. "A description of each action's effects in each state"

Also, All MDPs match Q-Learning in the idea that all decisions must be independent of decisions made in the past [5].

[1] `https://towardsdatascience.com/simple-reinforcement-learning-q-learning-fcddc4b6fe56`
[2] `https://towardsdatascience.com/the-complete-reinforcement-learning-dictionary-e16230b7d24e`

For practical benefit, David Silver offers an example MDP. In Silver's MDP, the set of all

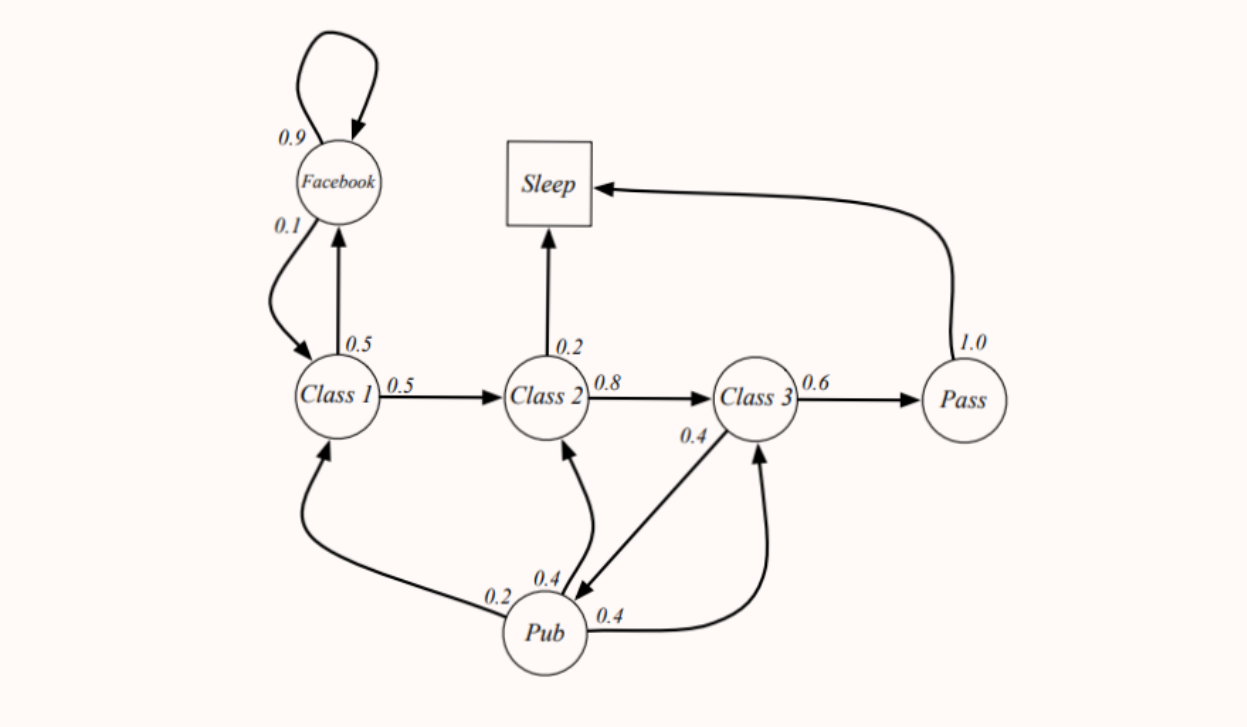


Figure 2.3: An MDP for Student Success Rate [5]

possible world states is all of the places a student could be: On Facebook, at the Pub, In any of the 3 classes, or at the "Pass state". The set of all possible actions is the collection of all *possible paths*. One path could be going to Classes 1 through 3 and then passing (which is our goal in this model). Alternatively, a student could go to the Pub, skip classes 1 and 2 and then go to class 3. The "real valued reward" in this case is how the decimal values present on the lines *accumulate*, giving higher reward weights to attending classes as opposed to visiting the pub. Step 4, the individual decimal values act as our "description of each action's effects in each state", however in other models, this could be accompanied by an *explanation of the action at hand* [37][5].

2.2.5 The Q-Equation and The Q-Learning Function

Understanding MDP is the foundation to understanding Q-Learning, as one of our objectives is to determine the best action in an MDP given a state in that MDP. However, the valued reward function described in the previous section is yet to be built. In Q-Learning, the process is to maximize the reward value at each iteration to achieve the optimal *"Q" or "Q-Function"*, using the *maximum Q-Value* at each step [38]. To start, we address a version of the *Bellman equation* [39][40]:

$$\boldsymbol{Q(s,a) = r + \gamma \max_{a'} Q\,(s', a')} \tag{2.1}$$

The basic version of this equation states that the Q value ($Q(s,a)$) for the current state (s) and current action (a) is determined by the current reward (r) added to the maximum future reward/Q-Value ($Q(\boldsymbol{s'}, \boldsymbol{a'})$) for the next state ($\boldsymbol{s'}$) and action ($\boldsymbol{a'}$) multiplied by a discount factor ($\boldsymbol{\gamma}$). The discount factor exists to ensure our algorithm prioritizes "...the discounted future reward at every step [39]."

This basic equation expands to the full *Q-Learning Equation* for determining reward values/Q-Values [39]:

$$\boldsymbol{Q_{t+1}}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right) = \boldsymbol{Q_t}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right) + \boldsymbol{\alpha}\left(\boldsymbol{r_{t+1}} + \boldsymbol{\gamma} \max_{\boldsymbol{a}} \boldsymbol{Q_t}\left(\boldsymbol{s_{t+1}}, \boldsymbol{a}\right) - \boldsymbol{Q_t}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)\right) \tag{2.2}$$

This is the same equation as before, with two modifications. The first is that this equation addresses the next Q-value ($\boldsymbol{Q_{t+1}}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)$) with respect to our current value Q-value ($\boldsymbol{Q_t}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)$). The second modification is the inclusion of the alpha ($\boldsymbol{\alpha}$), which represents "...the learning rate that controls how much the difference between previous and [the new] Q value is considered". [39] Given enough increments, our model will produce the *maximal Q-equation* which determines the *best action at every state* [39]. To reiterate, provided a state in an MDP, the *Q-Equation* will produce the best action from that state to get closer to the goal.

Symbol	Meaning
$\boldsymbol{Q_{t+1}}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)$	The next Q-Value
$\boldsymbol{Q_t}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)$	The current Q-Value
$\boldsymbol{s_t}$	The current state
$\boldsymbol{a_t}$	The current action
$\boldsymbol{t}$	The current step
$\boldsymbol{t+1}$	The next step
$\boldsymbol{\alpha}$	The learning rate, which determines the weight of the difference between the current value and the next
$\boldsymbol{r_{t+1}}$	The next reward value
$\boldsymbol{\max_a Q_t}\left(\boldsymbol{s_{t+1}}, \boldsymbol{a}\right) - \boldsymbol{Q_t}\left(\boldsymbol{s_t}, \boldsymbol{a_t}\right)$	The maximal next Q-Value
γ	A discount factor to ensure that a future value is always selected

Table 2.3: The Symbols of the Q-Learning equation

2.2.6 Q-Learning in the Context of Diagnosing Sepsis

For the sake of understanding the specific medical applications in Q-Learning, Yu C. et. al have surveyed the many different types of research taking place currently, all of which are using some form of reinforcement learning, including variations of Q-Learning [41]. In another study specific to Sepsis, Gottesman et. al discuss guidelines to be used by reinforcement

learning algorithms in the diagnosis of Sepsis, with a careful note to maintain "..caution and due diligence.." in implementation, as these are working with human lives [42]. Their discussion revolves mostly around the considerations that need to be made when researching with reinforcement algorithms. They also note the importance of providing as many details about a patient as possible to build a model, while also warning of confounding factors, or factors that affect one another. Q-Learning has been used already to diagnose Sepsis in patients, as seen in the Komorowski et. al study in detail later [27]. With the possibility of solving an MDP with two binary states/goals: *the patient does or does not have Sepsis*; Q-Learning has the potential to diagnose Sepsis.

2.3 Causal Inference

Normally, Q-Learning relies on the creation of an MDP before solving. Sometimes, the individual designing an algorithm can spot and create an MDP naturally based on the context of the model, like in the case of the previous example *Figure 2.3*. In the case of diagnosing an illness, often the actions that lead from one diagnosis state to another are unclear. To avoid the random ordering and testing of MDPs, we can employ a technique known as *Causal Inference* to add a methodology behind optimization.

2.3.1 Statistical Inference

Within the past sixty years, statistical inference has dominated in many fields, including Computer Science, as the primary means of determining the significance of resulting data from an experiment [43]. In particular, statistical inference is used to "...[compare] particular statistics from one observational data set to another ... with an appropriate reference distribution to *judge the significance of those statistics* [44]." At its basis, the goal of statistical inference is to determine if there exists a relationship between some data, whether that relationship means anything, and how the knowledge of that relationship might translate to the theoretical "whole set of data" known as the population [45]. For the problem of diagnosing Sepsis, we want to determine relationships among the available patient data, and create a model that represents certain characteristics of a given patient's data (like their blood pressure, body temperature, etc) as *predictors*, to determine if that patient has Sepsis or not [46]. This is done to build a model that represents the whole population, in this case, Sepsis patients.

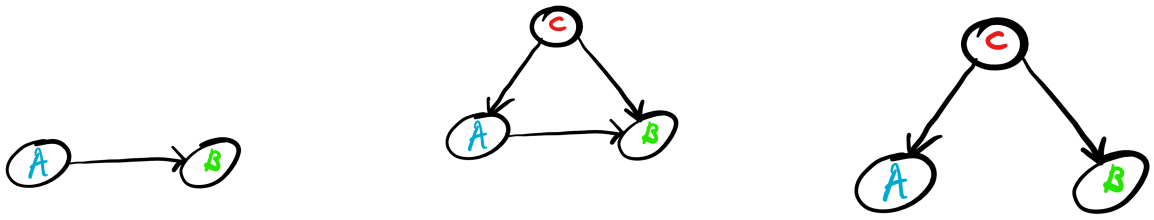
2.3.2 The Development of Causal Inference

The science of statistical inference lacks one particular ability that no model in its domain can address. Statistical inference cannot address the "*Why?*" behind relationships among data. Seeking to put an end to the idea that "correlation does not imply causation" prevents any understanding of causation through statistical means, Judea Pearl has spent the better part of the past three decades developing a field of statistics known as "*Causal Inference*",

netting a Turing award in 2012 [47][48]. As opposed to the traditional understanding of statistical inference, Pearl notes that Causal Inference opts to apply science methodology to understanding the cause of effects, as opposed to "...leaving causal consideration to the mercy of intuition and good judgement..." [49]. The science of Causal Inference seeks to claim that "...causal conclusions drawn from a carefully designed experiment are often valid" [50].

2.3.3 Building a Model in Causal Inference

Journalist Marin Pogančić offers a small example to illustrate the purpose of causal inference. Suppose we would like to determine, for a sample of individuals, the relationship between a person's chocolate consumption and their likelihood to win a Nobel Prize. Our initial causal diagram, provided by Pogančić would appear as *Figure 2.4a*. Node A represents



(a) Stage 1 (b) Stage 2 (c) Stage 3

Figure 2.4: The Simple Causal Model [6]

a person's chocolate consumption and Node B represents a person's Nobel Prize holding status (yes or no). From this basic model, there may be some known common cause that the modeler is aware of that impacts both Node A and Node B (such as country of origin). This common cause Node is deemed C in the new diagram, *Figure 2.4b*. With the introduction of the *Reichenbach Common Clause Principle*, which states "...if variables A and B have a common cause C, then when we condition on C, the correlation between [A and B] is wiped out..." [6][51]. The final diagram, shown in *Figure 2.4c* offers a clearer understanding that the causal relationship does not lie between A and B (chocolate and Nobel status), but rather between C and A and C and B(country and chocolate and country and Nobel status) [6].

2.3.4 Building More Complex Models

While the previous example does offer a simplified model, there are cases where certain predictors are the work of much more complex causal relationships. In the primer on the topic, "Causal Inference in Statistics", Pearl J. et al. describe three distinct pieces to building more complex models, known as *Structural Causal Models (SCMs)* [52]:

1. *Chains*
2. *Forks*
3. *Colliders*

An SCM *chain* is any causal relation shown from one *factor* (node) to another. In the case of the previous example, *Figure 2.4a* is a chain of two elements, from factor A (chocolate consumption) to factor B (Nobel Prize status). However, can be any number of factors long. An SCM *fork* represents a causal relationship from one factor to two or more different factors. *Figure 2.4c* from the previous example is also an example of a fork, which demonstrates a causal relationship from factor C (country of origin) to factor A, and factor C to factor B. The final form of causal relationship in an SCM is the SCM *collider* which demonstrates a causal relationship of two or more factors causing a single other factor. By flipping the arrows in *Figure 2.4c*, a collider can be shown with a causal relationships from A to C and B to C [47].

2.3.5 Understanding SCM Scope

SCMs are a method of demonstrating and building models within the lens of causal inference, but much like every other statistical method, some assumptions need to be addressed. A first concern is that the model builder is responsible for designating *exogenous variables* which are deemed outside the scope of the model, and *endogenous variables* which are inside the scope of the model. If a study is looking at the impact of burning various chemicals on the global environment, the researchers would deem several variables worth modeling as endogenous: temperature, population density, wildlife population. On the other hand, there are likely several variables that are not worth observing: GDP for an area, yearly fruit production. This *does not mean* that these factors do not have a causal relationship with a factor in the model, it simply means that the researcher has excluded them for the sake of scope [47].

2.3.6 Intervention

Particularly useful for medical applications, there are cases where through experimental trial runs, researchers may wish to inject "*Interventions*", where the researchers or something else modifies the conditions during the experiment to achieve the desired outcome [53][54][55]. In some cases, "...where randomized controlled experiments are not practical...", particularly in the cases of recorded data, intervention is still possible [47]. In an SCM, we can intervene by taking any given factor and fixing its value. Intervention offers the model-builder or experimenter the chance to substitute known values for values that a model would determine based on its findings. These are particularly useful when it is well known what impact a particular factor may have on an output. Fixing different factors' values and testing would produce models of varying quality, allowing a researcher to potentially achieve a better overall model with intervention.

2.3.7 Applications for Sepsis Diagnosis

The modeling for Q-Learning, MDP, and the modeling for causal inference, are very similar in structure and objective. As such, the information one gains from one model could be used to the benefit of the other and vice versa [56][57][58]. Taking it one step further, it seems likely that an MDP model could be constructed from a causal inference model. With the correct constraints applied, this would enable experimentation with Q-Learning that would provide insight into not only the impact of predictors on a result but also the impact of the predictors on each other. In the context of diagnosing Sepsis, causal inference would provide several models that would explain how changes in one of a *patient's vitals affect another of the same patient's vitals*. A *more accurate MDP* would lead to a *more accurate Q-Learning model solution*. A more accurate Q-Learning model solution would lead to *a more accurate diagnosis of Sepsis*.

2.4 Related Work

The inspiration for this paper, as addressed previously, is based on the 2019 "*Early Prediction of Sepsis from Clinical Data ... Computing in Cardiology Challenge*" provided by PhysioNet [28]. This challenge provided entrants with a database of various patient data, including whether or not a patient had Sepsis or not. The goal was to predict if a patient had Sepsis, 6 hours before the diagnosis from an actual doctor. During the judging phase, there was a much larger database of more patient data and diagnoses that the entrant's model was run against. The model gained points for every successful early diagnosis and lost varying amounts of points for late or incorrect diagnoses [59]. Many of these contestants utilized a form of Machine Learning, including the winning team which used a "...signature-based regression model" that took advantage of the techniques of both "...supervised and unsupervised machine learning models [60]."

The other central inspiration for this paper, also mentioned already, was research presented in a Nature article conducted by Komorowski et. al. The approach used by this team was an implementation of a Q-Learning algorithm as available in authors' GitHub repository [27][3]. The AI clinician was tasked with a similar objective as the PsyioNet challenge models: provide an optimal treatment strategy for a given Sepsis patient. Provided patient data, the AI clinician would supply an optimal treatment regiment by "...[extracting] implicit knowledge from an amount of patient data that exceeds by many-fold the life-time experience of human clinicians" in order to recommend a treatment [27]. Although the intent of this particular study focuses on treatment as opposed to diagnosis, the approach to the problem is similar.

Raghu et. al tackled the Sepsis treatment problem using a what is known as "Double Deep Q-Learning" [61]. Their approach attempts to minimize the loss between the output of their model, and the target at every state. The "double" portion of this kind of network comes from the main Q-Learning network sending output Q-Values using a "feed-forward"

[3] `https://github.com/matthieukomorowski/AI_Clinician`

pass, instead of having these values being determined directly from the target network. This kind of Q-Learning also makes use of what is called a "Dueling Q Network" which goes against the traditional model of each state being assigned a Q-Value for each given action. The Dueling Q Network splits the action value pair into an additional parameter called "advantage", which determines the quality of the best action at a given state. In terms of policy choice, the researchers opted not to prescribe vasopressors (medicine that makes blood pressure go up) unless the patient had a very high SOFA score. The researcher's model ended up performing very well at diagnosing and treating patients with medium-level SOFA scores(5 - 15), but there was a significant increase in mortality rates for patients with very high SOFA scores($>$15). The researchers concluded that the model could be recommended for mid-level SOFA scores, but not for high-level SOFA scores due to lack of data.

In another reinforcement learning paper by Raghu et. al, they were able to reduce patient mortality from Sepsis by up to 3.6% from the baseline mortality of 13.7% [62]. This model makes use of the previous hybridization of the "Double Deep Q-Learning Network" and the "Dueling Q Network" mentioned previously, with the addition of a new optimal policy technique. The "Doubly Robust Off-policy Value Evaluation" created by Jiang and Li is mentioned here as the means of gathering a given Q-Value at a particular point. This paper also draws comparisons between two of the authors' created techniques: the "Dueling Double-Deep Q Network (DDQN)" and the "Sparse Autoencoder Dueling DDQN". In their experiment, both models developed good practices for prescribing vasopressors over IV fluids, but the autoencoder variant was much more accurate at prescribing IV fluids (better than the human physician).

Peng et. al follow in the footsteps of the Raghu et. al group with another attempt using the DDQN method. The difference in this experiment is the blend of the DDQN approach with what is described as the Kernel Reinforcement Learning approach (KRL) [63]. The KRL approach groups nearby "neighbor states" and makes a distributed actions based on the closest neighbor states. Peng et. al proceeds to engage in what they describe as a "Mixture of Experts (MOE)" approach where they blend features from both the DDQN approach and the KRL approach. Also, they focused on optimizing a select set of factors, including age, Elixhauser (a medical index), SOFA, FiO2, BUN (Blood Nitrogen), Glasgow Coma Scale, Albumin, trajectory length, and max distance from neighboring states. In the results, the MOE approach behaved closest to the KRL approach but was benefited by the higher dosage nature of the DDQN approach. The authors conclude that the MOE approach outperforms the individual experts, as well as the clinician policy, with a note that more testing would need to be conducted on larger data sets to prevent extreme responses in outlier cases in particular.

Petersen et. al attempted a different route bearing the same idea in using Deep Q-Learning [64]. The authors of this experiment generated an MDP that deals with various cytokine groups in the human body. Cytokines are crucial in cell-signaling and are a way of knowing if something is wrong in the human body. The research group built their Deep Q-Learning model to treat areas where cytokine groups were producing active signals. Rewards were given to the model for a healed patient, and punishments for patients that died. The

model was able to obtain 0% mortality on the data set it was trained on, and achieves 0.8% mortality on 500 randomly selected "patient parameterizations" that had baseline mortality values between 1% and 99%. The authors note that in this experiment, the model medicates cytokines for which no known drugs are known to be capable of doing, likely allowing for such high output values.

Tsoukalas et. al chose to solve a Partially Observable Markov Decision Process (POMDP) using an algorithm known as Perseus to build a Clinical Decision Support System (CDSS) for Sepsis.[65] The purpose of a CDSS is to give a clinician an assistant tool to recommend actions based on patients' conditions, which are determined by the Perseus solution to the created POMPD. When following the optimal policy, the resultant policy enabled 25.9% of patients to transition to a better state 90% of the time, while 33.7% of patients transition to a worse state 90% of the time. This is an improvement to the non-policy cases which were 12.9% and 51.2% respectively.

RESEARCH GROUP	*MODEL TYPE*
PhysioNet Winner	Signature-Based Regression Model Mix of Supervised + Unsupervised
Komorowski et. al AI Clinician	Stock Q-Learning from MatLab
Raghu et. al	Dueling Double Deep Q-Learning
Raghu et. al Optimization	DDQN and Sparse Autoencoder DDQN
Peng et. al	MOE (DDQN + KRL)
Petersen et. al	Stock Deep Q-Learning on Cytokines
Tsoukalas et. al	POMDP + Perseus Algorithm

Figure 2.5: A Summary of the Related Work

Chapter 3

Research Overview

3.1 Proposed Methodology

3.1.1 Introduction to the Methodology

The entirety of this research process can be summarized in three steps:

1. Gather and construct a full collection of data
2. Use Q-Learning on the full collection of data to produce an optimal model
3. Further optimize that model with Causal Inference techniques

However, the intricacies of those steps need to be addressed at full, starting with a discussion on where exactly all the patient data will be sourced from.

3.1.2 Gathering PhysioNet Sepsis Patient Data

The PhysioNet Computing in Cardiology Challenge 2019 provided all applicants two data sets, training set A and training set B, containing information on 20,336 subjects and 20,000 subjects respectively.[1] This data was gathered from 3 hospitals in the past 10 years, from three different EMR systems (Electronic Medical Record) systems at Beth Israel Deaconess Medical Center (Boston, Massachusetts), Emory University Hospital (Druid Hills, Georgia), and a third unidentified hospital system [66]. Information on each patient at varying time intervals throughout a day are available, alongside 40 other factors that contain a patient's vitals (Heart Rate, O2 Saturation, Temperature) as well as whether or not a patient was determined to have Sepsis at a given point in time. Since this challenge has concluded, the two datasets are now publicly available and can be used for their intended purpose of building Sepsis prediction models. This data set does not contain entries for all patients at all times, neither does it contain a patient's full set of vitals at every given possible point. All of that being said, the data set is well defined and structured and will provide a basis for training the initial Q-Learning model.

[1] Challenge Data can be found here: `https://archive.physionet.org/users/shared/challenge-2019/`

3.1.3 Constructing MIMIC-III Data Set

In addition to the problem challenge data set, PhysioNet, through the MIT Lab for Computational Physiology, provides one of the largest publicly accessible intensive care databases in the form of *MIMIC-III*. MIMIC-III covers over a decade of medical records, and offers free and open access to anyone who has obtained "CITI 'Data or Specimens Only Research'" class completion.[2] In particular, the data covers over a decade of patient medical data, chart information, vitals data, and diagnosis information [67]. This amounts to close to 17,083 cases of varying stages of Sepsis. Due to the work of researchers who have come before this point, MIMIC-III has been made much more manageable for coming up with a Sepsis specific solution. Komoroski et. al, built a full MatLab script[3] that extracts patient information from the MIMIC-III data set. The result is a significantly more refined data set containing patients diagnosed with some form of Sepsis, as well as their vitals information, mirroring the content of the two challenge data sets. Komoroski et al. in addition provides a script to extract all the Sepsis-III definition patients and clean their values that are out of range or not correct in some way. This will provide a large body of information on the kinds of patients that test positive for Sepsis. Additionally, unlike the PhysioNet data set which is concretely timestamped throughout a patient's stay, the MIMIC-III data set does not offer quite this luxury. They offer two different, but useful groups of information used for model building.

3.1.4 Building the Initial MDPs

Before any machine learning is conducting, a Markov Decision Process needs to be created that represents the medical diagnosis process. The AI Clinician from Komorowski et. al's work uses a clustering technique known as "k-means++" to partition groups of patients based on criteria from their time-series data [27][68]. The Komoroski et. al research group has provided their MDP generator which forms 500 distinct MDPs from the MIMIC-III data set.[4] Their approach makes use of generating the 500 MDPs, training them on the MIMIC-III dataset, determining the best policy out of the 500 policies, and then testing that MDP and model combination on another data set. There are several ways to generate MDPs for diagnosis, including random selection of states and the actions between those states; however, it makes more sense to utilize a chosen data set to build states that are reflective of that data. The generation of large samples of MDPs that are directly reflective of the input data set will likely prove more useful than random selection. In particular, the end goal of all these MDPs will be the binary diagnosis of Sepsis (yes or no).

[2]MIMIC-III can be found here: `https://mimic.physionet.org/`

[3]Source Code can be found here: `https://github.com/matthieukomorowski/AI_Clinician/blob/master/AIClinician_mimic3_dataset_160219.m`

[4]The Script can be found here: `https://github.com/matthieukomorowski/AI_Clinician/blob/master/AIClinician_core_160219.m`

3.1.5 Implementing the Q-Learning Algorithm

In order to have a Q-Learning solution for the MDPs generated in the previous step, the Markov Decision Processes Toolbox is one of the offerings from the MathWorks community that has a full set of tools to perform Q-Learning without the necessity of implementation from scratch.[5] This is the route that the Komorowski research group opts to use, but should this pose a challenge for any reason, there are complimentary offerings in Python as well.[6] Both of these tools offer a succinct way to execute Q-Learning to solve an MDP with a single method call with desired parameters.

3.1.6 Training Initial Q-Learning Models

With data gathered, several MDPs to choose from, and a Q-Learning toolkit available to run, all that is needed is to train the model to solve the MDP and validate the outcome. This is where the input data discussed earlier comes into play. Operating only in the context of Training Sets A and B as well as the MIMIC-III dataset, the experiment lends itself into the following permutations:

- Train with Training Set A or Set B, Validate with MIMIC-III
- Train with Training Set A and B together, Validate with MIMIC-III
- Train with MIMIC-III, Validate with Training Set A or B
- Train with MIMIC-III, Validate with Training Set A and B together
- Perform Cross-Fold Validation on the combination of the two data sets

However, in order to train in this manner, MIMIC-III's labels/columns need to match the Training Sets A and B. As this likely will require a large amount of overhead, for now, the experiment will proceed entirely on MIMIC-III. The training and validation phases performed on these models will provide a measure of how accurate each MDP and corresponding Q-Learning solution is at attempting to diagnosing Sepsis. At this point in the experiment, the best MDP and Q-Learning solutions can be selected for additional optimization.

3.1.7 Simplifying the Models by Omitting Factors

Before optimizing any model, it is useful to only include factors that have a significant impact on the output (which in this case is the diagnosis for a patient). As such, this phase of the experiment is done to pare down factors in the best-performing models to see if they can be made simpler. Simpler models in statistics tend to be preferred as a general trend

[5] A link to the MDP Toolbox: `https://www.mathworks.com/matlabcentral/fileexchange/25786-markov-decision-processes-mdp-toolbox`

[6] Link to the MDP Toolbox for Python `https://pymdptoolbox.readthedocs.io/en/latest/api/mdptoolbox.html`

[69]. All of this with the goal of producing a model that remains applicable for other data sets that are not the input set. The best MDPs from the previous step will have a few of their factors removed, retested with Q-Learning to determine evaluation quality, and then will be kept or dismissed based on performance. Models whose simpler version demonstrate equitable quality in diagnosis Sepsis will be preferred and kept over their more complicating counterparts. This process will be repeated for all the top achieving models until they begin suffering a severe quality in diagnostic capability, at which point the last, best models will be selected to finalize the experiment. This kind of simplification is often implemented in what is known as 'feature selection'.

As for hyper-parameter optimization,

3.1.8 Optimizing the Models with Causal Inference

Removing factors is a way to make a model simpler, but it does not offer the same degree of manipulation that a Causal Inference approach does. Making use of the some known factors to help diagnose Sepsis discussed in the background to this paper (low blood pressure, high respiratory rate), may provide an experimental way to the output models. The results of the Q-Learning will provide the best actions at every given state, but modifying the MDPs to route a specific set of states and actions in a particular offer, may produce a greater accuracy rating. In addition to modifying the order of the individual states and the actions between them to support a medical approach to the problem, fixing factors that are common across all the best MDPs may prove additionally useful at determining how much reward needs to be placed taking certain actions at certain points in the decision chain. Much like the simplification step, the best MDPs will go through several rounds of rearranging actions and states, until there is no intuitive reasoning for modifying ordering or fixing any values. The experiment will conclude after the preceding steps have been fulfilled in the order in which they are laid out. Once the final round of Causal optimization has taken place, the best of all available models will be selected as the optimal experimental model, and the MDP that represents it will be deemed as the most optimal for the experimental environment.

3.2 Evaluation Methodology & Metrics

The "output" for this particular experiment is a finalized model that, when provided information on a patient, can determine if that patient has Sepsis or not. At each stage of building and testing the model, there is an evaluation conducted in the form of cross-validation with the second test data set, whichever one that ends up being. However, the overall quality of the specific output model that results from this experiment can be compared to the existing literature. Since this experimental design seeks to replicate that of the Komorowski et. al experiment, all evaluation will be taken with consideration of the results of their AI Clinician. In addition, Nemati et. al developed a model using Machine Learning practices that observes about 85% accuracy in diagnosing Sepsis with ICU related data in predicting Sepsis, providing a good benchmark figure to aim towards [70]. In addition, the standard

Machine Learning metrics of recall, precision, F1, and ROC will be taken into consideration [7].

3.3 Dissemination of Results

Ultimately, this experimental process will yield a thesis that contributes to the overall knowledge of the academic community as a whole. As such, it is necessary that it is open and available to those who may be able to benefit and learn from its content. The following are a list of three potential journal venues for the adoption of this publication:

- ACM Transactions on Modeling and Computer Simulation (TOMACS)[8]
- Simulation: Transactions of the Society for Modeling and Simulation International[9]
- Frontiers in Artificial Intelligence[10]
- Journal of Medical Internet Research (JMIR)
- BMC Medical Research Methodology
- International Journal of Medical Informatics
- BMC Medical Informatics and Decision Making
- EPJ Data Science
- ACM Transactions on Data Science

All of these are very simulation and modeling driven publications that accept papers from a wide domain of subject matter, including medical solutions. They are all also very machine-learning driven, and the kinds of papers that appear in these journals are in line with what the format that the result of this experiment should take.

[7] `https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c`
[8] Found here: `https://dl.acm.org/journal/tomacs`
[9] Found here: `https://journals.sagepub.com/home/sim`
[10] Found here: `https://www.frontiersin.org/journals/artificial-intelligence`

3.4 Timeline of Research

Month	Description
May 2020	Propose Thesis
June 2020	N/A
July 2020	Construct Full Data Set Set Up Full Database Begin Running Data Filtering Scripts
August 2020	Finish Up Running Data Cleansing Scripts Position, Format, and Build Initial MDPs
September 2020	Set Up and Implement Q-Learning Scripts Begin Training, Solving Initial MDPs with Varied Data Sets
October 2020	Finish Q-Learning Runs Select Optimal MDPs Begin Culling Excess Factors
November 2020	Continue Culling Excess Factors Rerun Q-Learning on New MDPs Create New Models
December 2020	Finish Culling Phase Finalize Set of MDPs and Desired Training Sets Continue to Optimization Phase
January 2021	Optimize MDPs Rerun Q-Learning on New MDPs
February 2021	Finalize Optimization Select Final Optimal Model Tidy Output, Data
March 2021	Write Thesis
April 2021	Write Thesis
May 2021	Defend Thesis
June 2021	Submit to Journal Publication

Chapter 4

Conclusion

This proposal lays out the foundation by which the experiment described within will take place. The result of the experiment will be an optimal Markov Decision Process, a Q-Learning approach to solve that MDP, and the novel contribution of optimizing it by use and way of Causal Inference techniques. Furthermore, the model produced at the end will provide some indicator of the feasibility of this kind of technique for Sepsis diagnosis. If the model ends up being weaker than some of its academic counterparts, this will likely prompt discussion on further ways that a different approach can be used to modification and optimization. If the resultant model offers significant promise, a similar approach can likely be applied to the diagnosis of other seemingly ambiguous conditions, as well as providing further optimization for treatment-based approaches to these kind of problems. In general, the results of this study are intended to further the simulation and modeling community's understanding of how reinforcement-based machine learning approaches can enable quicker diagnoses and the saving of more lives overall.

4.1 Future Work

This thesis proposal is intended to support the launchpad for the development of a full thesis. The final product, the finished thesis, will discuss the full methodology of the experiment in whole. The final thesis will include many of the elements that are on display in this paper (a background, related work) with the addition of all the crucial parts of the experimental process itself (the data gathering, the method, the results). Once the work is finished, the goal is that the final thesis will be used in part or whole to benefit other researchers to gain additional footing in the field of Reinforcement Learning based prediction.

Appendix A

Additional Tables and Figures

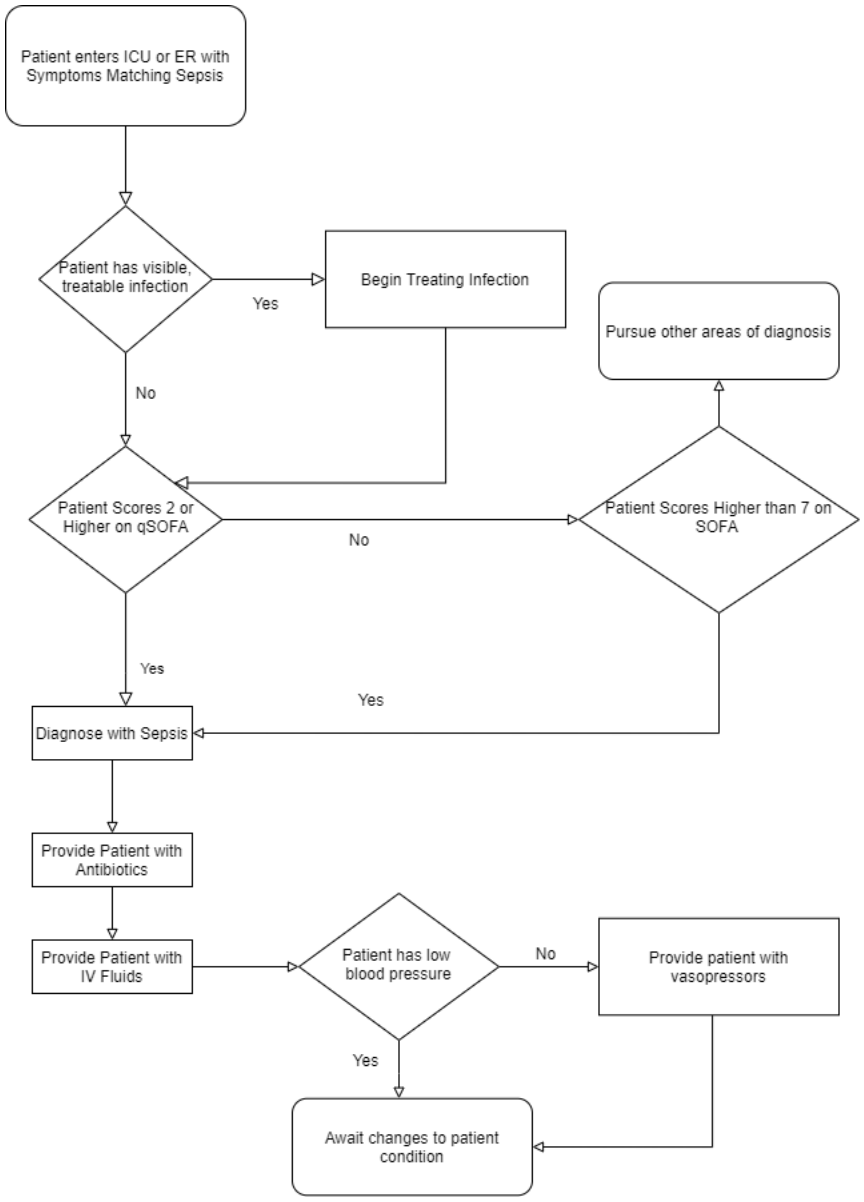


Figure A.1: A Flowchart for Understanding Sepsis Treatment and Diagnosis

References

[1] Krista O'Connell. Sepsis: Symptoms, causes, treatment, risks more, Aug 2018.

[2] Sharath Nair, Anilkumar Surendran, Rajmohan Prabhakar, and Meer Chisthi. Comparison between FOUR score and GCS in assessing patients with traumatic head injury: a tertiary centre study. *International Surgery Journal*, 4:656, 01 2017.

[3] CRISMA Center, University of Pittsburgh, and UPMC. qSOFA: quick Sepsis Related Organ Failure Assessment.

[4] Duffy Susan and Lifespan. Leading the Way in Treatment for Sepsis. *Emergency Services at Hasbro Children's Hospital*, 2020.

[5] David Silver. Markov Decision Processes. *UCL - London's Global University*, Dec 2018.

[6] Marin Vlastelica Pogančić. Causal vs. Statistical Inference, Jun 2019.

[7] Brad Brown, Michael Chui, and James Manyika. Are You Ready for the Era of 'Big Data'? *McKinsey Quarterly*, 2020.

[8] Andrew M. Nunn, Justin S. Hatchimonji, Daniel N. Holena, Mark J. Seamon, Brian P. Smith, Lewis J. Kaplan, Niels D. Martin, Patrick M. Reilly, C. William Schwab, and Jose L. Pascual. Boarding ICU patients: Are Our Rounding Practices Subpar? *The American Journal of Surgery*, 215(4):669–674, April 2018.

[9] K. Doi. Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. *Comput Med Imaging Graph*, 31(4-5):198–211, 2007.

[10] M. Singer, C. S. Deutschman, C. W. Seymour, M. Shankar-Hari, D. Annane, M. Bauer, R. Bellomo, G. R. Bernard, J. D. Chiche, C. M. Coopersmith, R. S. Hotchkiss, M. M. Levy, J. C. Marshall, G. S. Martin, S. M. Opal, G. D. Rubenfeld, T. van der Poll, J. L. Vincent, and D. C. Angus. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8):801–810, Feb 2016.

[11] CDC. What is Sepsis? *Centers for Disease Control and Prevention*, Aug 2019.

[12] Kristina E. Rudd, Sarah Charlotte Johnson, Kareha M. Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V. Colombara, Kevin S. Ikuta, Niranjan

Kissoon, Simon Finfer, Carolin Fleischmann-Struzek, Flavia R. Machado, Konrad K. Reinhart, Kathryn Rowan, Christopher W. Seymour, R. Scott Watson, T. Eoin West, Fatima Marinho, Simon I. Hay, Rafael Lozano, Alan D. Lopez, Derek C. Angus, Christopher J. L. Murray, and Mohsen Naghavi. Global, Regional, and National Sepsis Incidence and Mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *The Lancet*, 395(10219):200–211, January 2020.

[13] Greg S Martin, David M Mannino, Stephanie Eaton, and Marc Moss. The Epidemiology of Sepsis in the United States from 1979 through 2000. *New England Journal of Medicine*, 348(16):1546–1554, 2003.

[14] MedlinePlus. Septic Shock. *MedlinePlus Medical Encyclopedia*, Mar 2020.

[15] F. Balloux and L. van Dorp. Q&A: What are Pathogens, and What Have They Done to and for Us? *BMC Biol.*, 15(1):91, 10 2017.

[16] Djillali Annane, Eric Bellissant, and Jean-Marc Cavaillon. Septic Shock. *The Lancet*, 365(9453):63–78, Jan 2005.

[17] M. H. Schoenberg, M. Weiss, and P. Radermacher. Outcome of Patients with Sepsis and Septic Shock after ICU Treatment. *Langenbeck's Archives of Surgery*, 383(1):44–48, Mar 1998.

[18] Martin Fraser Clark and Simon Victor Baudouin. A Systematic Review of the Quality of Genetic Association Studies in Human Sepsis. *Intensive care medicine*, 32(11):1706–1712, 2006.

[19] WebMD and Khatri Minesh. What is Sepsis or Septicemia (Blood Infection)? *WebMD*, May 2019.

[20] John C. Marshall. *66 - Scoring Systems for Sepsis and the Multiple Organ Dysfunction Syndrome*, page 921–931. Academic Press, Jan 2001.

[21] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7):707–710, Jul 1996.

[22] J. R. Banegas, J. J. de la Cruz, F. Rodriguez-Artalejo, A. Graciani, P. Guallar-Castillin, and R. Herruzo. Systolic vs Diastolic Blood Pressure: Community Burden and Impact on Blood Pressure Staging. *J Hum Hypertens*, 16(3):163–167, Mar 2002.

[23] Royal College of Physicians and Surgeons of Glasgow. Assessment of Glasgow Coma Scale. *THE GLASGOW STRUCTURED APPROACH to ASSESSMENT of the GLASGOW COMA SCALE*.

[24] R. L. Gauer. Early Recognition and Management of Sepsis in Adults: the First Six Hours. *Am Fam Physician*, 88(1):44–53, Jul 2013.

[25] A. Kumar, D. Roberts, K. E. Wood, B. Light, J. E. Parrillo, S. Sharma, R. Suppes, D. Feinstein, S. Zanotti, L. Taiberg, D. Gurka, A. Kumar, and M. Cheang. Duration of Hypotension before Initiation of Effective Antimicrobial Therapy is the Critical Determinant of Survival in Human Septic Shock. *Crit. Care Med.*, 34(6):1589–1596, Jun 2006.

[26] CDC. Do You Know How Sepsis Is Diagnosed and Treated? *Centers for Disease Control and Prevention*, Jul 2019.

[27] Matthieu Komorowski, Leo A. Celi, Omar Badawi, Anthony C. Gordon, and A. Aldo Faisal. The Artificial Intelligence Clinician Learns Optimal Treatment Strategies for Sepsis in Intensive Care. *Nature Medicine*, 24(11):1716–1720, October 2018.

[28] Matthew Reyna and Gari Clifford. Early Prediction of Sepsis from Clinical Data – the PhysioNet Computing in Cardiology Challenge 2019, 2019.

[29] Stephen Marsland. *Machine Learning: an Algorithmic Perspective*. Chapman Hall/CRC machine learning pattern recognition series. CRC Press, second edition edition, 2015.

[30] Doina Precup, Richard S. Sutton, and Satinder Singh. Eligibility Traces for Off-Policy Policy Evaluation. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 759–766. Morgan Kaufmann, 2000.

[31] A. E. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford. Machine Learning and Decision Support in Critical Care. *Proc IEEE Inst Electr Electron Eng*, 104(2):444–466, Feb 2016.

[32] J. H. Chen and S. M. Asch. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N. Engl. J. Med.*, 376(26):2507–2509, Jun 2017.

[33] Zhongheng Zhang and written on behalf of AME Big-Data Clinical Trial Collaborative Group. Reinforcement learning in clinical medicine: a method to optimize dynamic treatment regime over time. *Annals of Translational Medicine*, 7(14), 2019.

[34] Yuan Ling, Sadid A. Hasan, Vivek Datla, Ashequl Qadir, Kathy Lee, Joey Liu, and Oladimeji Farri. Learning to Diagnose: Assimilating Clinical Narratives using Deep Reinforcement Learning. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 895–905, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.

[35] Christopher John Cornish Hellaby Watkins. *Learning from Delayed Rewards*. PhD thesis, King's College, Cambridge, UK, May 1989.

[36] Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, May 1992.

[37] Bob Givan and Ron Parr. An Introduction to Markov Decision Processes. *Exploration of Large State Spaces*, Nov 2001.

[38] P. Mehta and S. Meyn. Q-learning and pontryagin's minimum principle. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 3598–3605, 2009.

[39] Venelin Valkov. Solving an MDP with Q-Learning from scratch — Deep Reinforcement Learning for Hackers (Part 1), Apr 2019.

[40] Avinash K. Dixit. *Optimization in Economic Theory*. Oxford Univ. Press, 2. ed, paperback ed. repr. 1997 edition, 1997.

[41] Chao Yu, Jiming Liu, and Shamim Nemati. Reinforcement Learning in Healthcare: A Survey. 2019.

[42] Omer Gottesman, Fredrik Johansson, Matthieu Komorowski, Aldo Faisal, David Sontag, Finale Doshi-Velez, and Leo Anthony Celi. Guidelines for Reinforcement Learning in Healthcare. *Nature Medicine*, 25(1):16–18, 2019.

[43] Bradley Efron. Trevor Hastie. *Computer Age Statistical Inference*. Cambridge University Press, 2016.

[44] Pat Bartlein. Geographic Data Analysis. *University of Oregon*, 2018.

[45] R. E. Kass. Statistical Inference: The Big Picture. *Stat Sci*, 26(1):1–9, Feb 2011.

[46] Subrata Kumar Das. *Computational Business Analytics*. Chapman Hall/CRC, 2014.

[47] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, first edition edition, 2018.

[48] Bob Brown. Judea Pearl, a Big Brain Behind Artificial Intelligence, Wins Turing Award, Mar 2012.

[49] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2000.

[50] Paul W. Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[51] Hans Reichenbach and Maria Reichenbach. *The Direction Of Time*. Dover books on physics. Dover, 1999.

[52] Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell. *Causal Inference In Statistics: a Primer*. John Wiley Sons Ltd, 2016.

[53] S. A. Baldwin, E. Stice, and P. Rohde. Statistical Analysis of Group-Administered Intervention Data: Reanalysis of Two Randomized Trials. *Psychother Res*, 18(4):365–376, Jul 2008.

[54] P. E. Lekone and B. F. Finkenst?dt. Statistical Inference in a Stochastic Epidemic SEIR Model with Control Intervention: Ebola as a Case Study. *Biometrics*, 62(4):1170–1177, Dec 2006.

[55] Ken'ichi Morooka. A Survey on Statistical Modeling and Machine Learning Approaches to Computer Assisted Medical Intervention: Intraoperative Anatomy Modeling and Optimization of Interventional Procedures. *IEICE Transactions on Information and Systems*, E96-D:pp.784–797, 04 2013.

[56] Seng-Beng Ho. Causal Learning Versus Reinforcement Learning for Knowledge Learning and Problem Solving. 2017.

[57] Ferenc Huszar. Causal Inference 3: Counterfactuals, Jan 2019.

[58] Fredrik D. Johansson, Uri Shalit, and David Sontag. Learning Representations for Counterfactual Inference. 2016.

[59] Matthew A. Reyna, Christopher S. Josef, Russell Jeter, Supreeth P. Shashikumar, M. Brandon Westover, Shamim Nemati, Gari D. Clifford, and Ashish Sharma. Early Prediction of Sepsis From Clinical Data. *Critical Care Medicine*, 48(2):210–217, February 2020.

[60] J. Morrill, A. Kormilitzin, A. Nevado-Holgado, S. Swaminathan, S. Howison, and T. Lyons. The Signature-Based Model for Early Detection of Sepsis From Electronic Health Records in the Intensive Care Unit. In *2019 Computing in Cardiology (CinC)*, pages Page 1–Page 4, 2019.

[61] Aniruddh Raghu, Matthieu Komorowski, Imran Ahmed, Leo A. Celi, Peter Szolovits, and Marzyeh Ghassemi. Deep reinforcement learning for sepsis treatment. *CoRR*, abs/1711.09602, 2017.

[62] Aniruddh Raghu, Matthieu Komorowski, Leo Anthony Celi, Peter Szolovits, and Marzyeh Ghassemi. Continuous state-space models for optimal sepsis treatment - a deep reinforcement learning approach. *CoRR*, abs/1705.08422, 2017.

[63] X. Peng, Y. Ding, D. Wihl, O. Gottesman, M. Komorowski, L. H. Lehman, A. Ross, A. Faisal, and F. Doshi-Velez. Improving Sepsis Treatment Strategies by Combining Deep and Kernel-Based Reinforcement Learning. *AMIA Annu Symp Proc*, 2018:887–896, 2018.

[64] Brenden K. Petersen, Jiachen Yang, Will S. Grathwohl, Chase Cockrell, Claudio Santiago, Gary An, and Daniel M. Faissol. Precision Medicine as a Control Problem: Using Simulation and Deep Reinforcement Learning to Discover Adaptive, Personalized Multi-Cytokine Therapy for Sepsis. *CoRR*, abs/1802.10440, 2018.

[65] Athanasios Tsoukalas, Timothy Albertson, and Ilias Tagkopoulos. From Data to Optimal Decision Making: A Data-Driven, Probabilistic Machine Learning Approach to Decision Support for Patients With Sepsis". *JMIR Med Inform*, 3(1):e11, Feb 2015.

[66] Matthew A. Reyna, Christopher S. Josef, Russell Jeter, Supreeth P. Shashikumar, M. Brandon Westover, Shamim Nemati, Gari D. Clifford, and Ashish Sharma. Early prediction of sepsis from clinical data: The physionet/computing in cardiology challenge 2019. *Critical Care Medicine*, 48(2), 2020.

[67] Alistair E.W. Johnson, Tom J. Pollard, Lu Shen, Li wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3(1), May 2016.

[68] David Arthur and Sergei Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, USA, 2007. Society for Industrial and Applied Mathematics.

[69] K. Jaqaman and G. Danuser. Linking Data to Models: Data Regression. *Nat. Rev. Mol. Cell Biol.*, 7(11):813–819, 11 2006.

[70] S. Nemati, A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit. Care Med.*, 46(4):547–553, 04 2018.