© Copyright 2021

Andrew K. Teng

Acute care sepsis prediction:
Analyzing the influence of social and behavioral determinants

Andrew K. Teng

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Adam B. Wilcox, Chair

Anne M. Turner

David Carlbom

Program Authorized to Offer Degree:

Biomedical and Health Informatics

University of Washington

Abstract

Acute care sepsis prediction:
Analyzing the influence of social and behavioral determinants

Andrew K. Teng

Chair of the Supervisory Committee:
Professor Adam B. Wilcox
Biomedical and Health Informatics

Within the hospital setting, sepsis is a leading cause of mortality, affecting more than 1.7 million adults annually. It is also present in about 30 to 50 percent of hospitalizations that end with death. Despite the high incidence and prevalence, detection and diagnosis of sepsis remain a challenge due to its non-specific early stage symptoms. However, as it can quickly progress to a life-threatening stage, it is important to detect sepsis patients earlier to improve outcomes. With the recently increased adoption of EHRs, many institutions now have large amounts of patient

data being collected and have created their own customized sepsis detection and mortality tools using various modeling or machine learning (ML) techniques. Additionally, those who experience more socioeconomic challenges are more susceptible to chronic illnesses, including sepsis. However, structured coding of social or behavioral features is often underutilized and unreliable. First, in order to understand the current environment of predictive analytics solutions for sepsis, we systematically identified various studies that utilize different models or ML techniques and analyzed their approach and results. Second, we developed a framework that utilizes natural language processing text classification from clinical notes to extract social and behavioral determinants of health (SBDH). Third, we assessed classification methods that utilize currently established sepsis definitions or clinical scores to establish a baseline and integrated the SBDH data extracted from clinical notes described earlier and determined if SBDH features can help enhance predictive performance for sepsis detection in the acute care setting.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee members for their support and guidance. A special thank you to my committee chair and advisor, Adam Wilcox, for his counsel and ability to direct me to anyone who could provide insight and assistance if he was unable to do so himself. I would also like to thank David Carlbom for taking a chance and meeting me a few years back, along with Rosemary Grant. Our initial meetings on sepsis and acute care inspired this work and I am fortunate to have been able to work with you. I am grateful to Anne Turner for initially being my academic advisor and having someone to check-in with during my start into the program. From your public health informatics class, I was able to gain insight and a desire to explore social determinants in the realm of population health. Additionally, to the Biomedical and Health Informatics department, faculty and staff, thank you for providing the advisement, funding, and resources necessary to prosper.

Next, I'd like to thank the great people I have met in this program and am glad to be able to call them my friends. From my cohort specifically, Calvin Apodaca and Jason Thomas, we have endured this journey together and although our outcomes vary slightly, it has been a great ride. Another special thanks to Jim Phuong, Reggie Casanova-Perez, and Wesley Surento for keeping me sane during the pandemic. A shoutout as well to Yifan Wu, Hannah Burkhart, Harsh Patel, and Carolin Spice for being a part of the food and karaoke squad.

Lastly, I want to thank my mom and dad, Sara and Kevin, for being my main cheerleaders, encouraging me to aim high, and pursue this doctoral degree. Your unwavering love and support have shaped me into being the achiever that I am today, and this journey would not be possible without you. Thank you.

Dedicated to my mom and dad.

Chapter 1. INTRODUCTION

1.1 SIGNIFICANCE OF THE PROBLEM

Sepsis is a leading cause of mortality in hospitals nationwide; the Centers for Disease Control and Prevention (CDC) reports about one third of patients who died in a hospital, died from sepsis.[1,2] Even though the occurrence and prevalence of sepsis is fairly high, there are many challenges with early detection and diagnosis of sepsis as it typically presents with early nondescript onset symptoms that are common to other illnesses, such as elevated body temperature and high heart rate.[3] Because sepsis can progress to life-threatening stages in patients rapidly, it is a priority for many to be able to identity and treat patients prior to the progression to increase recovery and survival outcomes. Additionally, the high prevalence and cost disease of sepsis is financially straining on healthcare institutions. In 2013, around $24 billion was spent on care for patients with sepsis, making it the most expensive condition for U.S. hospitals.[1] Increased risk for sepsis is also associated with other chronic diseases such as chronic obstructive pulmonary disease, cirrhosis, and other immunodeficiency disorders.[4] Furthermore, those who survive sepsis may also suffer from additional long-term complications, such as higher risk of hospital readmission, increased cardiovascular disease severity, and potentially cognitive impairment, all of which may lead to reduced life expectancy and/or lower quality of life.[5] With the Health Information Technology for Economic and Clinical Health (HITECH) Act, Meaningful Use, and Institute for Healthcare Improvement (IHI) Triple Aim framework, the adoption of electronic health records (EHRs) and other supporting technologies increased with aims to improve healthcare quality, patient safety, and data security.[6–9] With this increased adoption of EHRs, secondary analysis at healthcare institutions is now feasible due to the large amounts of data being collected and stored.

1.2 CURRENT SEPSIS DEFINITIONS

It has been challenging to define sepsis and various definitions have been created over the years. In 1991, the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference introduced the first clinical definition of sepsis, severe sepsis, and septic shock, referred as Sepsis-1.[10] Sepsis-1 defined sepsis as a systemic inflammatory response syndrome (SIRS) due to a present infection, with at least two abnormal vital sign criteria.[11,12] Additionally, Sepsis-1 aggregated the definition of sepsis further, creating severe sepsis and septic shock.[11,12] Because SIRS is nonspecific, the presence of at least two SIRS criteria may not always be the result of an infection. Therefore, in 2001, the International Sepsis Definitions Conference updated Sepsis-1 with the revised definition of Sepsis-2, which introduced confirmed or suspected infection to the prior sepsis definition.[11,13] However, another update to the definition was introduced in 2016, Sepsis-3, and now defines sepsis as a life-threatening organ dysfunction caused by a dysregulated host to infection.[12] There are currently two scoring schemes available for determining organ dysfunction: (1) Sequential Organ Failure Assessment (SOFA) Score, which is used to calculate organ dysfunction related to sepsis, often used in the ICU setting, and (2) quickSOFA (qSOFA), which is used to identify patients with high risk of adverse outcomes, often used in the non-ICU setting.[12] SOFA takes various organ systems into account, including the respiratory, hematologic, hepatic, cardiovascular, neurologic, and renal systems.[12] On the other hand, qSOFA criteria consists of vital signs and mental status.[12]

Outside the United States, detection and defining sepsis is also challenging. In 1999, the Audit Commission in London recommended developing early warning systems (EWSs) to help clinical staff identify patients who needed attention. They proposed a modification of Morgan's Early Warning Score, creating the Modified Early Warning Score (MEWS) which aimed to create communication between nursing and medical staff when patient deterioration was detected.[14] MEWS consists of the following clinical features: (1) systolic blood pressure, (2) heart rate, (3) respiratory rate, (4) temperature, and (5) Alert-Verbal-Pain-Unresponsive (AVPU) score. These concrete flow-based definitions have been applied in the clinical setting to assist with sepsis detection with varying success.[15–18]

1.3 IMPACT OF TECHNOLOGY

With increased EHR adoption and usage in recent years, many institutions have started to utilize the vast amount of collected patient data and create customized sepsis detection and mortality tools using machine learning (ML) techniques. At a high level, ML is the application of algorithms and/or artificial intelligence to aid with automatic learning and detection. Although there are many ML algorithms, they are commonly categorized as supervised or unsupervised methods. Supervised ML algorithms take what has been learned through the use of a labeled training data set to create a model that is then applied to a new unlabeled dataset, often referred to as a testing or validation set, to predict future outcomes or events.[19] Because supervised methods assume truth from labeled data, if the labeled data are noisy or incorrect, the resulting predictions may not be reliable, as the generated models are trained on these potentially incorrect data. On the other hand, while unsupervised ML algorithms draw inferences and patterns from datasets to describe unlabeled data, commonly used in data exploration as they can highlight potential trends in data without existing data analysis.[19] Both of these techniques can be used on structured or unstructured data. Structured data typically encompasses quantitative data, such as lab result values or vital sign values. However, unstructured data, such as clinical text notes from providers or images, can provide an array of information not present or identifiable in structured data.[20] Using natural language processing (NLP) to analyze notes can potentially provide greater insight as not all patient data is readily available in structured formats.

ML methods have been used to aid with sepsis prediction. Locally at Harborview Medical Center (HMC), a systematic rule-based screening tool was developed in 2012 to identify septic inpatients using EHR data to evaluate patients for signs of infection. The screening system utilized SIRS criteria as well as lactate and would notify providers if nurses suspected new or worsening infection. Internal pilot testing revealed that rapid response team activations were reduced by half on the acute care floor. Additionally, it was shown that time from sepsis screen to care was reduced and 3-hour bundle compliance increased over a year. Furthermore, the average nurse response time to screen decreased.[21,22]

However, with this increased adoption, there have been challenges standardization and data interoperability. Sharing data between institutions, and sometimes within institutions, can be challenging depending on how EHRs are implemented. Additionally, there are many EHR platforms which structure and store data differently, and each institution may even customize their systems even further to meet their criteria and workflow. However, collating and extracting these data can be challenging due to these nuances in how the data are collected and stored. Prior to March 2020, HMC had contracts with two major EHR vendors, Epic and Cerner, where Epic was used primarily in the outpatient setting (e.g. neighborhood clinics, emergency department) while Cerner was used in the inpatient setting (e.g. intensive care, emergency department transfer). Each of these two vendors collected and stored patient data differently, which added hurdles for data sharing and interoperability. However, using a common data model (CDM), such as the Observational Medical Outcomes Partnership (OMOP) CDM, would allow for data from different databases and EHRs to be combined seamlessly and transparently, ultimately adding to reproducibility and ease of use.[23,24] Data analysis can then be performed on the CDM, creating code and scripts that would be applicable to other data that are transformed to the same common data model, increasing its utility and generalizability.[24–26]

1.4 SOCIAL AND HEALTH BEHAVIOR DATA

In addition to structured data and commonly used vital signs and lab results, social and behavioral determinants of health (SBDH) data are now starting to be analyzed for secondary research as recent research has indicated that there is a correlation between SBDH and health outcomes.[27] Common SBDH indicators can not only include specifically social determinants such as housing stability, access jobs and health care services, education level, language, and socioeconomic conditions, but also health behaviors, such as diet, physical activity, alcohol intake, and tobacco use.[28–30] These SBDH indicators are descriptors of populations since they are useful predictors of health outcomes and health related interventions.[31] Because SBDH can potentially be strong indicators of health, many institutions are now starting to analyze and intake SBDH information, whether through clinical text notes or standardized coding, such as International Classification of Diseases (ICD).[30] However, there are challenges with SBDH intake as there is no standardized SBDH screening tool in the EHR; additionally, coding schemes like ICD can prove to be unreliable in secondary analysis as coding can oversimplify symptoms and diagnoses leading to coding uncertainties and the fact that coding errors may be present from unintentional mistakes or even upcoding.[32,33] Furthermore, SBDH relies heavily on the clinical staff to screen and document SBDH as well as assumes patients will provide accurate answers. Complications arise when documenting diagnostic codes and SBDH.[32] Past research has shown that hospital readmissions are highly influenced by patient health status and SBDH and suggest that clinical staff and researchers should consider SBDH when assessing readmission risk.[34]

Housing stability is a major public health issue. Within our region, it is estimated that there are at least 22,000 individuals facing housing insecurity in Washington State, with more than 12,000 in Seattle/King County, a four percent increase over the previous year.[35] Housing instability is associated with various health inequalities, such as shorter life expectancy, higher morbidity, and increased usage of acute hospital services, "as the social determinants of homelessness and health inequities are often intertwined, and long term homelessness further exacerbates poor health".[36] Although some research has shown that patients who experience housing instability are more likely to die following admission for severe sepsis than those with insurance,[37] other research indicates that the effects of health inequalities are still unclear and need further investigation.[38] Additionally, social habits and health behavior, such as tobacco and alcohol use, can impact health decisions and outcomes; one study found that participants who drank alcohol and reported tobacco use consumed more foods higher in fat and sugar, low in vitamins and minerals as well as foods, considered to be less healthy and prepared in a less healthy way.[39] Therefore, it may be important to look at SBDH and their impact on health outcomes.[39]

1.5 SEPSIS AND SOCIAL AND HEALTH BEHAVIORAL DATA

The 2018-2019 King County Community Health Needs Assessment (CHNA) found local healthcare issues, including affordability and stability, mental health, tobacco, and alcohol and substance-related disorders, that were major challenges affecting our community.[40] Past research has shown that low socioeconomic status, including low income or housing stability issues, contributes to an increased mortality and intensive care unit (ICU) admission in patients with sepsis.[41] Sepsis has also been found to be associated with demographic and social factors, such as tobacco and alcohol use, poverty level, and race; additionally, access to care is closely correlated with lower sepsis prevalence and mortality rates.[4]

When evaluating the impact of homelessness on presentation and outcomes, it was found that homeless sepsis patients were younger and had more liver disease.[42] These patients also used more healthcare resources, longer length of stays, increased risk of dying within 30 days of admission, and tended to show higher 180-day mortality.[41,42] Similarly, a study from Argentina that explored health inequities in the diagnosis and outcome of sepsis found that patients in socially disadvantaged groups were sicker at admission, had septic shock more frequently, and had higher mortality.[43]

Additionally, smoking and tobacco use are leading causes of preventable illness and death; they are often associated with the development of chronic diseases and conditions, such as respiratory difficulties, lung cancer, bronchitis and bacterial pneumonia.[44–46] Smoking can increase susceptibility to other bacterial infections which can lead to septic shock and respiratory failure.[44] Recent literature has also found that smoking is associated with longer hospital stays, increased need for mechanical ventilation, and higher mortality.[45,47] On the other hand, it has also been shown that current smoking is associated with a decreased risk of mortality in pneumococcal pneumonia with bacteremia.[46] Given the complex nature of the effect of smoking on immune function, it is difficult to predict the overall impact of tobacco smoking on clinical outcomes in sepsis and shows that this social feature is worth exploring. Similarly, alcohol is frequently abused, and it has been shown that there is a disproportionate level of alcohol-related consequences among those who classify as low socioeconomic status.[48] Chronic alcohol consumption can impair a person’s antibacterial defense against infections; therefore, patients with chronic alcohol consumptions or experiencing alcohol use disorders can become more severely ill and are at an increased risk of developing infections that can lead to sepsis.[49,50] This is due to higher cortisol levels, which have been found to be elevated in patients with septic shock, and decreased immune systems.[51]

Overall, social and behavioral factors have been shown to impact health outcomes and sepsis onset. Although some research has started to show that patients who experience health inequalities like housing instability are more likely to die following admission for severe sepsis, other research indicates that the effects of health inequalities are still unclear and need further investigation. Therefore, it is important to explore how social and behavioral health determinants can impact health outcomes.

1.6 STATEMENT OF THE STUDY PURPOSE

Although there have been inroads in exploring how predictive analytics can be used for sepsis prediction, most studies have focused on patients in the intensive care unit (ICU) or emergency department (ED) and most have focused on applying SIRS, Sepsis-2 or Sepsis-3 definitions as gold standards. However, to our knowledge, little has been investigated in the acute care setting; which we define as consisting of patients who have been admitted to the hospital for a stay that is greater than 24 hours and do not require critical care. This research will attempt to discover a new method that integrates SBDH for sepsis prediction or detection in the acute care setting at HMC. Through use of supervised algorithmic-based text and sentiment classification ML methods on clinical text notes, we hope to shed new light on potential social features, which are collected in unstructured formats at varying points in a patient’s stay, that may have a meaningful impact on predictive analytics for sepsis patients in acute care. Despite all these varying approaches to adapting predictive analytics for sepsis, there has not been an instance to our knowledge where social factors and past medical history are extracted from unstructured clinical text notes and transformed into structured metrics to be integrated into a predictive machine learning algorithm. There are three main objectives to this study:

Aim 1: Perform a systematic scoping review to understand the status quo of predictive analytics and sepsis.

Aim 2: Extract clinical text notes with social history, and other related features, and apply various supervised text classification algorithms to create a numeric scoring metric that captures sentiment and then compare the feasibility of extracting SBDH phenotypes using clinical text notes and structured disease coding.

Aim 3: Integrate the previously generated scoring metrics on social determinants, convert these metrics into features that will then be integrated with a general SIRS, Sepsis-3, and MEWS based classification sepsis prediction algorithms, evaluate their impact on prediction in the acute care setting and identify and analyze patterns that may arise through a mixed-methods approach.

1.7 CONTENT OF THE DISSERTATION

This dissertation discusses how SBDH data can be extracted from unstructured data sources and explores how integrating these derived data can impact sepsis prediction as defined by existing scoring systems commonly used in the acute care setting. Below is an outline of the content of each chapter:

In Chapter 2, we present findings from a systematic scoping review that we published in Applied Clinical Informatics.[52] Here, we explore the current literature and specifically examine the approaches and methods used in the literature to determine how others are defining sepsis in their respective patient population. Furthermore, we explore which algorithms or methods were used across studies to shape our decisions and methods for Aim 3.

In Chapter 3, we present findings from applying text classification methods on clinical notes to extract social and health behavior data, focusing specifically on (1) housing stability, (2) tobacco use, and (3) alcohol use as they were found to be locally important social and behavioral determinates. We also compare our findings from the unstructured classification with structured sources of SBDH data. We then weight and average the classifications for each patient by social factor to form a generalizable and interpretable risk score.

In Chapter 4, we present findings from integrating scores that were generated in Chapter 3 with clinical, rule-based, scoring schemes, focusing specifically on (1) qSOFA, (2) SIRS, and (3) MEWS. We analyze how the integration of SBDH data extracted from unstructured clinical notes can influence the predictability of these scoring methods.

In Chapter 5, we summarize our dissertation findings from all studies. Furthermore, we outline how our findings can be used to inform future work and discuss areas in where the research could be further explored.

1.8 REFERENCES FOR CHAPTER 1

1. O'Brien J. The Cost of Sepsis. The Centers for Disease Control and Prevention Blog. Published November 18, 2016. Accessed February 3, 2020. https://blogs.cdc.gov/safehealthcare/the-cost-of-sepsis/

2. Centers for Disease Control and Prevention. Data & Reports | Sepsis | CDC. Accessed August 11, 2019. https://www.cdc.gov/sepsis/datareports/index.html

3. Sepsis. Sepsis - Symptoms and causes - Mayo Clinic. Published November 16, 2018. https://www.mayoclinic.org/diseases-conditions/sepsis/symptoms-causes/syc-20351214

4. Reinhart K. *Sepsis: A Global Public Health Burden for Hospitals*.; 2018. https://www.beckershospitalreview.com/quality/sepsis-a-global-public-health-burden-for-hospitals.html

5. Deferio JJ, Breitinger S, Khullar D, Sheth A, Pathak J. Social determinants of health in mental health care and research: a case for greater inclusion. *J Am Med Inform Assoc*. 2019;26(8-9):895-899. doi:10.1093/jamia/ocz049

6. Health IT Legislation. HealthIT.gov. Accessed January 25, 2020. https://www.healthit.gov/topic/laws-regulation-and-policy/health-it-legislation

7. Anumula N, Sanelli PC. Meaningful Use. *AJNR Am J Neuroradiol*. 2012;33(8):1455-1457. doi:10.3174/ajnr.A3247

8. Henry J, Pylypchuk Y, Talisha Searcy, Patel V. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015. Published May 2016. https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php

9. The IHI Triple Aim. Insitute for Healthcare Improvement. http://www.ihi.org/Engage/Initiatives/TripleAim/Pages/default.aspx

10. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med*. 1992;20(6):864-874.

11. Balk RA. Systemic inflammatory response syndrome (SIRS): Where did it come from and is it still relevant today? *Virulence*. 2014;5(1):20-26. doi:10.4161/viru.27135

12. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801. doi:10.1001/jama.2016.0287

13. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med*. 2003;31(4):1250-1256. doi:10.1097/01.CCM.0000050454.01978.3B

14. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. *Ann R Coll Surg Engl*. 2006;88(6):571-575. doi:10.1308/003588406X130615

15. Kovach CP, Fletcher GS, Rudd KE, Grant RM, Carlbom DJ. Comparative prognostic accuracy of sepsis scores for hospital mortality in adults with suspected infection in non-ICU and ICU at an academic public hospital. Moreira J, ed. *PLOS ONE*. 2019;14(9):e0222563. doi:10.1371/journal.pone.0222563

16. Roney JK, Whitley BE, Long JD. Implementation of a MEWS-Sepsis screening tool: Transformational outcomes of a nurse-led evidence-based practice project. *Nurs Forum*. 2020;55(2):144-148. doi:10.1111/nuf.12408

17. Tirotta D, Gambacorta M, La Regina M, et al. Evaluation of the threshold value for the modified early warning score (MEWS) in medical septic patients: a secondary analysis of an Italian multicentric prospective cohort (SNOOPII study). *QJM*. Published online January 9, 2017:hcw229. doi:10.1093/qjmed/hcw229

18. Khwannimit B, Bhurayanontachai R, Vattanavanit V. Comparison of the accuracy of three early warning scores with SOFA score for predicting mortality in adult sepsis and septic shock patients admitted to intensive care unit. *Heart & Lung*. 2019;48(3):240-244. doi:10.1016/j.hrtlng.2019.02.005

19. Soni D. Supervised vs. Unsupervised Learning: Understanding the differences between the two main types of machine learning methods. Toward Data Science. Published March 22, 2018. Accessed March 5, 2020. https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d

20. Malmasi S, Hosomura N, Chang L-S, Brown CJ, Skentzos S, Turchin A. Extracting Healthcare Quality Information from Unstructured Data. *AMIA Annu Symp Proc*. 2017;2017:1243-1252.

21. Wolpaw BJ, Kress N, Grant RM, et al. Early Identification and Treatment of Hospital Acquired Sepsis: Success of an EHR-Based System. Presented at the: SGIM 2018.

22. Bean M. This Seattle hospital aims to get sepsis patients on antibiotics within 3 hours — here’s how. Becker’s Clinical Leadership and Infection Control. Published February 23, 2018. Accessed February 3, 2020. https://www.beckershospitalreview.com/quality/this-seattle-hospital-aims-to-get-sepsis-patients-on-antibiotics-within-3-hours-here-s-how.html

23. Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences*. 2016;113(27):7329-7336. doi:10.1073/pnas.1510502113

24. Hripcsak G, Duke JD, Shah NH, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform*. 2015;216:574-578.

25. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19(1):54-60. doi:10.1136/amiajnl-2011-000376

26. Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553-564. doi:10.1093/jamia/ocu023

27. Artiga S, Hinton E. *Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity*. KFF: Henry J Kaiser Family Foundation; 2018. Accessed February 7, 2020. http://files.kff.org/attachment/issue-brief-beyond-health-care

28. Social Determinants of Health. HealthyPeople.gov. Accessed February 1, 2020. https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health

29. Socio-Behavioral Determinants of Health (SBDH) Data Catalog. https://ictr.johnshopkins.edu/programs_resources/sbdh/

30. Tan M, Hatef E, Taghipour D, et al. Including Social and Behavioral Determinants in Predictive Models: Trends, Challenges, and Opportunities. *JMIR Med Inform*. 2020;8(9):e18084. doi:10.2196/18084

31. Social Determinants. Institute for Health Metrics and Evaluation. Accessed February 1, 2020. http://www.healthdata.org/social-determinants

32. Wockenfuss R, Frese T, Herrmann K, Claussnitzer M, Sandholzer H. Three- and four-digit ICD-10 is not a reliable classification system in primary care. *Scand J Prim Health Care*. 2009;27(3):131-136. doi:10.1080/02813430903072215

33. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620-1639. doi:10.1111/j.1475-6773.2005.00444.x

34. Lax Y, Martinez M, Brown NM. Social Determinants of Health and Hospital Readmission. *Pediatrics*. 2017;140(5):e20171427. doi:10.1542/peds.2017-1427

35. Henry M, Mahathey A, Morrill T, et al. *The 2018 Annual Homeless Assessment Report (AHAR) to Congress*. The U.S. Department of Housing and Urban Development OFFICE OF COMMUNITY PLANNING AND DEVELOPMENT; 2018. https://files.hudexchange.info/resources/documents/2018-AHAR-Part-1.pdf

36. Stafford A, Wood L. Tackling Health Disparities for People Who Are Homeless? Start with Social Determinants. *International Journal of Environmental Research and Public Health*. 2017;14(12):1535. doi:10.3390/ijerph14121535

37. Ahmad S, Baig S, Taneja A, Nanchal R, Kumar G. The Outcomes of Severe Sepsis in Homeless. *Chest*. 2014;146(4):230A. doi:10.1378/chest.1995140

38. Bambra C, Gibson M, Sowden A, Wright K, Whitehead M, Petticrew M. Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. *Journal of Epidemiology & Community Health*. 2010;64(4):284-291. doi:10.1136/jech.2008.082743

39. Papadopoulou S, Hassapidou M, Katsiki N, et al. Relationships Between Alcohol Consumption, Smoking Status and Food Habits in Greek Adolescents. Vascular Implications for the Future. *Current Vascular Pharmacology*. 2017;15(2):167-173. doi:10.2174/1570161114666161024123357

40. King County Community Health Needs Assessment 2018/2019. https://www.kingcounty.gov/depts/health/data/community-health-indicators/~/media/depts/health/data/documents/2018-2019-Joint-CHNA-Report.ashx

41. Schnegelsberg A, Mackenhauer J, Nibro HL, Dreyer P, Koch K, Kirkegaard H. Impact of socioeconomic status on mortality and unplanned readmission in septic intensive care unit patients. *Acta Anaesthesiol Scand*. 2016;60(4):465-475. doi:10.1111/aas.12644

42. Chiu C, Werge DA, Arab N, Palafox M, Minejima E. 1486. Evaluation of the Impact of Homelessness on Presentation and Outcomes of Gram-Negative Sepsis. *Open Forum Infectious Diseases*. 2019;6(Supplement_2):S541-S541. doi:10.1093/ofid/ofz360.1350

43. Estenssoro E, Loudet CI, Edul VSK, et al. Health inequities in the diagnosis and outcome of sepsis in Argentina: a prospective cohort study. *Crit Care*. 2019;23(1):250. doi:10.1186/s13054-019-2522-6

44. Bagaitkar J, Demuth DR, Scott DA. Tobacco use increases susceptibility to bacterial infection. *Tob Induced Dis*. 2008;4(1):12. doi:10.1186/1617-9625-4-12

45. Huttunen R, Laine J, Lumio J, Vuento R, Syrjänen J. Obesity and smoking are factors associated with poor prognosis in patients with bacteraemia. *BMC Infect Dis*. 2007;7(1):13. doi:10.1186/1471-2334-7-13

46. Beatty JA, Majumdar SR, Tyrrell GJ, Marrie TJ, Eurich DT. Current Smoking and Reduced Mortality in Bacteremic Pneumococcal Pneumonia. *Chest*. 2016;150(3):652-660. doi:10.1016/j.chest.2016.04.020

47. Alroumi F, Abdul Azim A, Kergo R, Lei Y, Dargin J. The impact of smoking on patient outcomes in severe sepsis and septic shock. *j intensive care*. 2018;6(1):42. doi:10.1186/s40560-018-0312-x

48. Collins SE. Associations Between Socioeconomic Factors and Alcohol Outcomes. *Alcohol Res*. 2016;38(1):83-94.

49. Gustot T, Fernandez J, Szabo G, et al. Sepsis in alcohol-related liver disease. *Journal of Hepatology*. 2017;67(5):1031-1050. doi:10.1016/j.jhep.2017.06.013

50. de Wit M, Wiaterek GK, Gray ND, et al. Relationship between alcohol use disorders, cortisol concentrations, and cytokine levels in patients with sepsis. *Crit Care*. 2010;14(6):R230. doi:10.1186/cc9385

51. Sam S, Corbridge TC, Mokhlesi B, Comellas AP, Molitch ME. Cortisol levels and mortality in severe sepsis. *Clin Endocrinol*. 2004;60(1):29-35. doi:10.1111/j.1365-2265.2004.01923.x

52. Teng AK, Wilcox AB. A Review of Predictive Analytics Solutions for Sepsis Patients. *Appl Clin Inform*. 2020;11(03):387-398. doi:10.1055/s-0040-1710525

Chapter 2. REVIEW OF PREDICTIVE ANALYTICS SOLUTIONS FOR SEPSIS PATIENTS

This is an Accepted Manuscript of an article published by Thieme in Applied Clinical Informatics on April 2, 2020, available online at https://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0040-1710525. © Georg Thieme Verlag KG

2.1 ABSTRACT

Background: Early detection and efficient management of sepsis are important for improving healthcare quality, effectiveness, and costs. Due to its high cost and prevalence, sepsis is a major focus area across institutions and many studies have emerged over the past years with different models or novel machine learning techniques in early detection of sepsis or potential mortality associated with sepsis.

Objective: To understand predictive analytics solutions for sepsis patients, either in early detection of onset or mortality.

Methods and Results: We performed a systematic scoping review and identified common and unique characteristics between their approaches and results in studies that used predictive analytics solutions for sepsis patients. After reviewing 148 retrieved papers, a total of 31 qualifying papers were analyzed with variances in model, including linear regression (n = 2), logistic regression (n = 5), support vector machines (n = 4), and Markov models (n = 4) as well as population (range: 24-198,833) and feature size (range: 2-285). Many of the studies used local datasets of varying sizes and locations while others used the publicly available MIMIC data. Additionally, vital signs or lab test results were commonly used as features for training and testing purposes; however, a few used more unique features including gene expression data from blood plasma and unstructured text and data from clinician notes.

Conclusion: Overall, we found variation in the domain of predictive analytics tools for septic patients, from feature and population size to choice of method or algorithm. There are still limitations in transferability and generalizability of the algorithms or methods used. However, it is evident that implementing predictive analytics tools are beneficial in the early detection of sepsis or death related to sepsis. Since most of these studies were retrospective, the translational value in the real-world setting in different wards should be further investigated.

Keywords: Sepsis, Predictive analytics, Machine learning, Algorithms, Data modeling

2.2 INTRODUCTION

Sepsis is a severe complication stemmed from an infection in the body and can lead to potential tissue damage, organ failure, or even death. More than 1.7 million individuals are diagnosed with sepsis annually in the United States and has a one in three mortality rate.[1] Sepsis is a strain on the hospitals and health care system as it is a disease of high prevalence and cost. In 2013, almost $24 billion was spent on care for sepsis patients, making it the most expensive condition to treat in U.S. hospitals.[2] Unfortunately, sepsis can stem from a vast array of initial infections, such as pneumonia or a urinary tract infection. Despite the high occurrence and prevalence, detection and diagnosis of sepsis remain a challenge due to its nondescript early onset symptoms, such as high heart rate and clammy skin.[3] However, as it can quickly progress to a life-threatening stage, it is crucial to treat sepsis patients earlier and more efficiently to increase survival outcomes. Furthermore, patients diagnosed with sepsis tend to remain in the hospital for a significantly longer periods of time when compared to those without the condition; thus, using more resources and hampering the ability to move patients out of the emergency department (ED) and into beds efficiently.

Currently, there are various metrics in use to define and identify sepsis in the clinical setting. In 1991, the Sepsis-1 definition of sepsis, severe sepsis, and septic shock was released. Sepsis was then described as a systemic inflammatory response syndrome (SIRS) due to a present infection, with at least two of the following criteria: (1) Temperature >38°C or < 36°C, (2) Heart rate > 90 bpm, (3) Respiratory rate > 20 or $PaCO_2$ < 32 mm Hg, or (4) WBC > 12,000/mm^3, < 4,000/mm^3, or > 10% bands; severe sepsis was having sepsis resulting in organ dysfunction while septic shock was the occurrence of sepsis-induced hypotension.[4] In 2001, an update resulted in the introduction of the Sepsis-2 definition, which added confirmed or suspected infection to the sepsis definition.[4] However, in 2016, Sepsis-3 was created and sepsis is now described as a life-threatening organ dysfunction caused by a dysregulated host to infection.[5]

While the definitions of sepsis have evolved, so has data collection in the clinical setting. In 2010, the United States government established a three-stage incentive program, aptly titled ‘Meaningful Use’, which established the requirement to use electronic health records (EHRs). With Meaningful Use stage one, EHRs were widely adopted and now streams of patient data are constantly being collected. Many researchers and clinicians are now trying to leverage and integrate the data to create tools that aid in early detection of sepsis. Many of these tools and predictive solutions use machine learning (ML) techniques or hazards model to assist in predicting sepsis onset or mortality. ML is the application of artificial intelligence to aid with automatic learning, detection, or classification, which potentially can be useful with medical data.[6] Additionally, the type and size of the feature set is important for the efficacy and interpretability of ML techniques as irrelevant features may lower the effectiveness.[7]

Because there are many ML models and feature sets that can be used for sepsis predictive analytics, we systematically identified various studies to understand the current state of sepsis prediction tools. Furthermore, we sought to determine how predictive analytics are being implemented for septic patients and to see if there are any optimal solutions for sepsis detection or mortality associated with sepsis currently being explored.

2.3 METHODS

We performed a literature search on PubMed in November 2018 to identify current practices and studies that have used predictive analytics for septic patients, aiding in both early detection of onset or mortality, using the following query: ("sepsis" OR "septicemia" OR "septic" OR "septic shock" OR "severe sepsis") AND ("prediction" OR "predict" OR "analytics") AND ("machine learning" OR "big data" OR "AI" OR "NLP" OR "neural network" OR "algorithm"). We included common synonyms and popular phrases for sepsis and predictive analytics to cast a greater net when searching. After querying the database, we identified and categorized relevant articles by reviewing the article titles and abstracts to ensure a ML technique or model was being used for prediction or detection of sepsis. Our search was not limited to the intensive care unit (ICU) as sepsis does not solely occur in such a setting. Furthermore, this allowed a larger scope to explore novel techniques being developed in other clinical settings. Our overall initial inclusion criteria are detailed below:

1. Study was published in a peer-reviewed journal or conference.
2. Study was published in English.
3. Study was published after 2008.
4. Study used at least one ML or model technique.
5. Study identified the features and dataset used.

6. Study presented their summary statistics and/or compared to previous studies.

Studies that did not meet these criteria were automatically excluded. However, to increase the breadth of our search and to include seminal studies, we added studies based on their titles and abstracts that were commonly found in the references, regardless of publication year. We did not restrict our search to a specific population age for similar reasons, but most of the reviewed studies centered around the general adult population aged 18 to 65. After curating our set of relevant articles, we identified and categorized the characteristics of each study by analyzing the common themes and differences between them to better understand the issues in applying predictive analytics to sepsis detection.

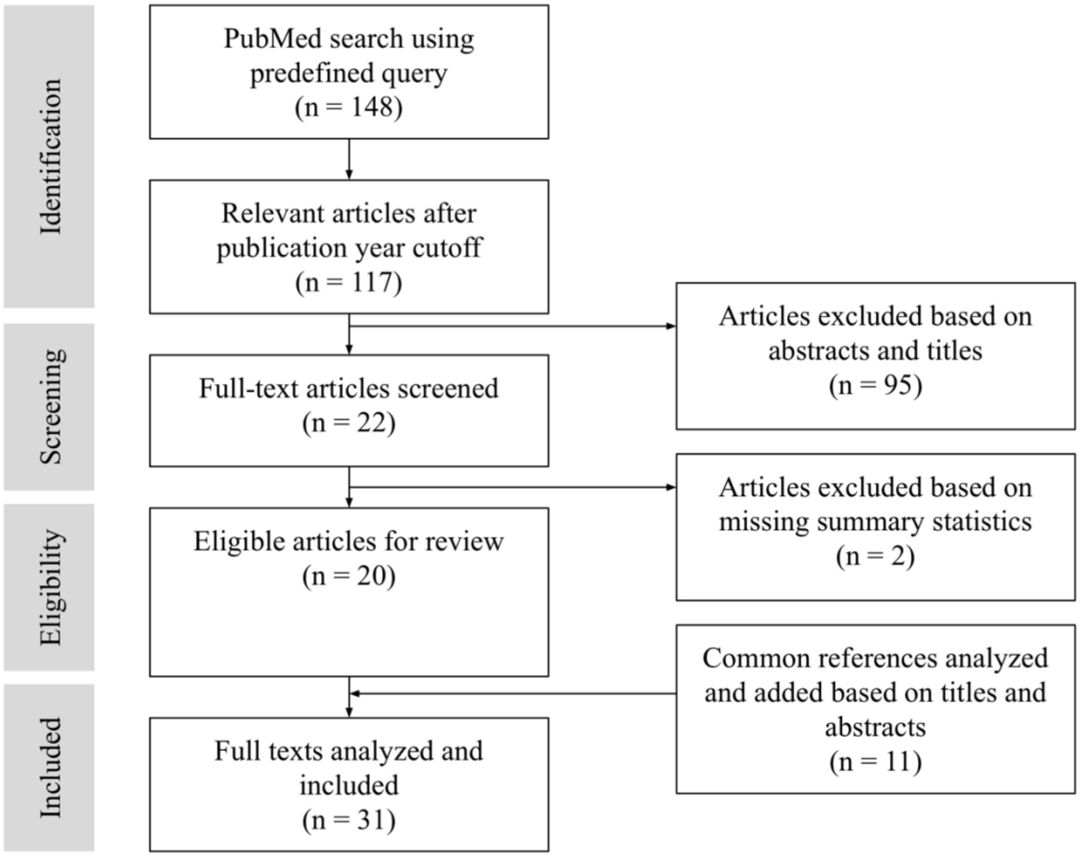


Figure 2.1. PRISMA adapted diagram.

2.4 RESULTS

Figure 2.1 depicts the article selection process. Our PubMed search resulted in 148 articles, where 31 articles were immediately excluded due to publication date. After reviewing the abstracts of the remaining articles, 95 articles were further excluded as they were not relevant to our question. A total of 22 full-text articles were extracted from our PubMed search. From these articles, a total of 11 cited references were added based on their titles and abstracts as they were commonly cited papers between many of the articles that met our inclusion criteria. However, two articles were then removed as there were no summary statistics presented. In total, 31 papers were analyzed. Full results are summarized in Table 2.1, detailing the algorithm or model, population and feature size, "gold standard" definitions, and summary statistics.

For the majority of the studies analyzed, a variety of metrics were used to report results, such as area under the receiver operator curve (AUROC) and accuracy. The reported metrics were dependent on the ML technique or model used, the features that were selected, and the size of the study data. From our review, we identified four key differences between the analyzed studies, (1) variability in ML or modeling techniques, (2) variability in feature selection, (3) variability in data sample selection and size and (4) variability in "gold standard" sepsis definitions.

2.4.1 *Variability in Machine Learning or Modelling Techniques*

A wide range of models and machine learning techniques were used to predict or detect sepsis onset, septic shock, severe sepsis, or mortality. The variety of methods used, summarized in Table 2.2, added to the richness of this systematic scoping review. Common methods include linear regression (n = 2), [8, 9] logistic regression (n = 5),[10-14] support vector machines (n = 4),[15-18] Markov models (n = 4),[19- 22] and Bayesian networks (n = 2).[23, 24] Additionally, a few studies (n =

6),[25-30] used an industry created tool, *InSight* (Dascena Inc.), to validate performance compared to the more commonly used methods. In particular, Mao et al., used *InSight* to test the predictive abilities of the industry created sepsis detection algorithm on open source and local datasets, determining the transferability of the algorithm across varying datasets.[28] Similarly, a few studies, including Danner et al., Gultepe et al., and, Thottakkara et al. used multiple methods and algorithms for comparison purposes against their own developed solution.[9, 15,17] Many of the general results concluded that utilizing predictive analytics were beneficial in the detection or prediction sepsis onset or mortality. Predictive performance measures for each study are detailed in Table 2.1. Due to heterogeneity in methods and standards used, the predictive performance measure varies across the analyzed studies. Table 2.1 details the goals of the studies, the "gold standard" or definition used for sepsis, septic shock, or severe sepsis, the best performance markers, and summary statistics.

2.4.2 *Variability in Feature Selection*

A majority chose common vital signs, including heart rate, temperature, respiratory rate, and diastolic and systolic blood pressures, for predicting sepsis onset, septic shock, severe sepsis, or mortality. However, some studies went beyond these common vital signs and found that including biological data in tandem with these common features could potentially help enhance prediction and detection. Sutherland et al. used blood cultured for gene expression analysis to help with their detection solution.[14] In addition to the common vital and test variables, they included 145 biomarkers to compare gene expression change from the Affymetrix Genechip data and were able to conclude that gene expression biomarker test had the ability to detect early evidence of sepsis before the availability of microbiology results. Langley et al. used blood and protein profiles to help curate individualized detection of sepsis.[13] Using these profiles, they found that patients with severe sepsis had more skewed distribution of metabolomic measurements and distinct metabolic differences between sepsis survivor and death groups. Afterwards, they created a solution that took clinical features and various metabolites to predict survival of patients with sepsis. Lukaszewski et al. created neural network models that would predict which ICU patients would develop sepsis from two daily samples of blood.[31] They used various leukocytes and cytokines (IL-1B, IL-6, Il-8, IL-10, TNF-a, CCL-2, Fas-L) as features for model development. However, they mentioned that the model may also be identifying individuals who are more likely to develop sepsis from a genetic predisposition. Although a few studies included biological data to aid with sepsis prediction and detection, it may not always provide better results. Stanculescu et al. used biological data from neonates for their real-time sepsis prediction tool.[22] They found that this addition was not statistically significant from their previous work using hidden Markov models and vital signs. From this, it is apparent that feature selection for machine learning techniques varies and there is not one set of features that is more ideal than another.

2.4.3 *Variability in Data Selection and Size*

Many studies used publicly available datasets, such as Medical Information Mart for Intensive Care (MIMIC) (n = 8)[8, 12, 19, 25-28, 32], or the less commonly used Medical Data Warehousing and Analysis (MEDAN) project (n = 2)[18, 33], to help train and validate their tools. The MIMIC-III dataset contains 53,423 distinct hospital admissions for adult patients (aged 16 years or above) admitted to critical care units between 2001 and 2012 while the MEDAN dataset contains data from 71 German ICUs from 1998 to 2002.[18,34] These datasets are extensive and provide researchers with real, de-identified data that can be used as testing, training or validation sets when using predictive analytics. Additionally, many studies (n = 23) used ICU data (either local[14,17,20-23,29-31,35-37] or MIMIC), while nine studies used ED[9-11,13-15,16,24,38,39] data. While local data varied greatly in size, ranging from 24 to 198,833, some used MIMIC in addition to their local datasets, which created a potentially more generalizable set of data to increase statistical significance and to increase the transfer of learning. Nemati et al. used local data as the development cohort and MIMIC data as the validation cohort rather than a random split of local data for both the development and validation cohorts.[35] This allowed them to claim that their solution is more generalizable and has the potential to work sufficiently well across institutions. Similarly, Mao et al. integrated both ICU and non-ICU using local mixed wards data to increase generalizability and MIMIC to increase transferability.[28]

Additionally, in our review, most of the studies reviewed were retrospective as they used data that was previously collected to create predictive analytics solutions, but there were four prospective studies in which tools were created to assist in the real-time clinical setting. Sutherland et al. prospectively predicted sepsis onset by using the American College of Chest Physicians/Society of Critical Care (ACCP/SCCM) consensus statement and if the patient had suspected infection based on microbiological diagnosis.[14] Using recursive partitioning, LASSO, and logistic regression on microarray procedures, they examined individual genes via a Bayes-adjusted linear model and leave-one-out cross validation. Later, they used 42 genes to generate a diagnostic classifier using a LogitBoost machine learning algorithm and applied the classifier to the validation set. Although the diagnosis of sepsis was unknown at the time of enrollment, confirmation was done retrospectively, and they found their real-time detection tool was able to perform before the availability of microbiology results.

Similarly, Lukaszewski et al. prospectively monitored molecular changes to identify presymptomatic individuals with an admission diagnosis of “likely septic”.[31] They used real-time

PCR to predict sepsis at an early stage of microbial infection, before overt clinical symptoms were to appear. Furthermore, they built five neural network classifiers, each with 30 percent of the data, to assess non-linear patterns and used a chi-squared test to ascertain whether the neural network derived predictive accuracies that were statistically significant. Although their solution was able to predict sepsis before the comparative method using the SIRS criteria, they found that clinicians might have trouble understanding the results from the neural network tool. Sawyer et al. pilot tested a real-time automated sepsis alert that would increase the rate of interventions within 12 hours of detection.[38] They found that their alert system resulted in an increase in early intervention for those who were identified to be at risk for sepsis.

2.4.4 *Variability in "Gold Standard" Definitions*

When implementing a retrospective predictive analytics solution, defining the outcome variable can greatly impact the performance. Among the analyzed studies, there was variability in defining sepsis. Some studies determined if a patient had sepsis by using the presence of an International Classification of Disease (ICD) Code 9 or 10, while others opted for a more rule-based approach based on the Sepsis-2, Sepsis-3, SIRS criteria, and/or organ dysfunction presence. Detailed information regarding gold standard definitions can be seen in Table 2.1. Additionally, a few studies relied on manual chart review for determining septic patients. On the other hand, prospective studies utilized a different approach. One utilized an admission diagnosis code upon ICU entry[31], while another study determined septic patients based on a real-time sepsis alert generated from their clinical alert system.[38]

2.5 DISCUSSION

In this study, we systematically reviewed the literature to identify all relevant studies that used a predictive analytics solution, including machine learning and hazards models, to predict onset or mortality of sepsis in hospitalized patients. We identified 31 studies and detailed the various methods and models each study utilized. Because the studies selected were not homogeneous in nature, there are a few distinct differences that should be noted.

Most of the studies reviewed were retrospective, while a few were prospective. Although many of the results show improved accuracy and early detection of sepsis onset or mortality, it remains unknown how effective and efficient many of these predictive analytics solutions are in a real-time patient care setting. To fully understand the usability and accuracy of these solutions, they should be studied prospectively and observed in the healthcare setting. Similarly, Michael found that prospective cohort studies could potentially capture clinically relevant variables that are absent from retrospective data sets and they could also gather data in a more representative an accurate manner.[40]

Furthermore, a lot of studies used ICU data, most likely due to data availability. The algorithms created using just ICU data may not be transferable to other departments due to the high variation in patient population, differences in scoring schemes, and possible missingness for features found in ICU data potentially being utilized in the predictive tool. For instance, often times organ dysfunction is measured different scoring schemes in the ICU and ED environments.[41] Although many studies presented favorable predictive value, it is nearly impossible to conclude from the reviewed studies if one predictive analytics solution is more effective than another as there were differences among gold standards. These gold standard definitions could have resulted in definition-specific results and if modified could yield differing results. Furthermore, the population and data sizes used for each study was different and the features that were used ranged from solely vital signs all the way to including free text and administrative data.[42] The heterogeneous nature of all the reviewed studies shows that there are many approaches for solving the question of applying predictive analytics for sepsis.

A few studies used industry created solutions that were sponsored by the respective company. These results can even be seen advertised on the company's websites themselves.[43] Because of this connection, there may have been publication bias present. It is best to be aware of this and interpret their respective results accordingly.[44]

2.5.1 *Limitations of Predictive Analytics Solutions Used*

Because of the heterogeneity of the data used, there were many differences in types of predictive analytic solutions used. Most commonly used predictive analytics techniques used linear classifiers, such as Naive Bayes and linear or linear and logistic regressions. Cross validation was also a common technique that was used amongst the analyzed studies. The use of cross validation indicates that the selected sample sizes in some of these explored studies may not be large enough. By utilizing cross validation, an artificial large sample size is created; however, by doing so, there is a risk of overfitting. Using k-fold cross validation can help reduce the effects of overfitting but does not eliminate the risk.[45] With machine learning, larger and more representative data sets can result in more realistic outcomes and higher predictive power. Therefore, it is important to consider the effects of data size. However, some models, such as basic linear regression, may oversimplify a real-world scenario as features and response variables may not follow a linear relationship. Multivariate linear regression can produce a more complete model in understanding the independent impact of predictor variables on an outcome;

similarly, multivariate logistic regression can only be used when the outcome variable is categorical, which may not always be the intended case.[46]

Non-linear models that were explored include neural networks and Markov models. Neural networks are easy to conceptualize, they are slower, do not have as great of performance metrics, requires tuning many parameters, and if a multi-layer neural network is used, then it is even harder to train.[47] Hidden Markov models were also considered, which are memoryless, and make assumptions that the next event is only dependent on the current event and not the past event. Markov models are state machines with the state changes being probabilities. In a hidden Markov models, the probabilities are not known, but the outcome is known. However, implementing a neural network may be too much of a black box and may not be ideal in a healthcare setting where doctors and clinicians would most likely want to be aware of the computations and reasoning behind the outputs.[48]

In addition to the common methods previously mentioned, there were a few unique methods that were used worth discussing. One study used Symbolic Gate Approximation which helped reduce the set of features necessary. Another study used Bayesian Principal Component Analysis (PCA), which can be advantageous for small data sets in "high dimensions as it can avoid the singularities associated with maximum likelihood PCA by suppressing unwanted degrees of freedom in the model".[49]

Generalized additive models were also used. These can be powerful in that they allow us to fit a nonlinear function to each predictor potentially allowing for more accurate predictions when compared to a linear method. Furthermore, because the model is additive, the effect of each predictor can be analyzed when fixing the remaining predictors; however, this additive characteristic can also be hindering.[50] One study used a Weibull-Cox proportional hazards model, which is a good method for analyzing survival data, and is smoother than just a standard Cox model. Finally, random forests were also explored. It is a bagging technique for both classification and regression. The general concept is that you divide your data into several portions, use a relatively weak classifier/regressor to process, and then combine them. Random forest is flexible and can enhance the accuracy of the weak algorithm to a better extent at the expense of heavier computational resources required.[51] However, if the data is not meaningful to begin with, the end result will still not be meaningful. While these uniquely applied predictive analytics solutions are interesting, they were not specifically differentiated from the more common tools that were used.[50] Overall, there was no clear-cut best algorithm; however, when selecting a predictive analytics solution to implement, one must consider the bias-variance tradeoff and sample size of the data.

2.5.2 *Limitations in Findings*

There are some limitations in terms of how the systematic scoping review was conducted and designed. We chose our search query to be fairly broad to be able to capture the variety of predictive analytics solutions being created for septic patients. Many of these studies had a different objective in their approach to using predictive analytics making it is difficult to determine whether one approach was better than another. If we narrowed the objective of our review to solely include studies that aimed to decrease sepsis mortality or onset, the search would have brought in different studies and examples. Furthermore, the definition of sepsis is ever changing, and this solely marks, generally, what has been done until now.

The availability of data seems to deeply affect and influence data availability. The most commonly used data were obtained from the ICU followed by the ED. This skew may be due to the availability of public open-access data, such as MIMIC and MEDAN. Therefore, we do not have good information in terms of whether or not predictive analytics tools are better applied in the ICU setting for better patient outcomes. As more data sets become available, we should be careful in interpreting where the application of these algorithms should be best assigned and used. The definition of sepsis that was used as well as target population contributed to the variation. Stanculescu et al. looked at developing an alert system for neonates. Because the number of neonates that met their eligibility criteria was low, their study population size was thus limited.[21, 22] Most of the studies included in our analysis attempted to detect sepsis or death related to sepsis earlier than what is currently available. Furthermore, most of the studies were retrospective, but there were a few prospective studies performed, and even a randomized control trial. When using machine learning to predict sepsis onset, many studies used vital signs or lab test results as their features to train and test their solutions. However, Sutherland et al. show that waiting for the microbiology results could potentially be avoided if gene expression analysis from blood plasma were to be utilized instead.[14] They were able to show strong findings of detecting sepsis before the availability of microbiology results. Shimabukuro et al. performed a randomized control trial by using an algorithm created by Dascena, Inc. and found that their predictor decreased the average length of stay and in-hospital mortality rate.[30] Nemati et. al found that they can predict, accurately, sepsis onset 4-12 hours prior to clinical recognition through the use of their modified hazards model.[35] Horng et al. found that utilizing the unstructured text data improved the accuracy of models that solely used the structured data.[16] However, since most of these studies were performed retrospectively, a prospective approach would be needed to determine the feasibility and clinical utility of these predictive analytics methods. For those that did use a prospective approach, there were varying results in accuracy metrics as some found that there was no major improvement in patient outcomes, while others found decreased in-hospital mortality and length of stay rates. Additionally, since the definitions of sepsis currently available rely on clinical features, bias in prediction models will be present as there will be an overlap in the feature set and outcome. Further research and exploration would be necessary in this area.

2.5.3 *Limitations in Search Strategy*

There are a few limitations in our search query. We used a simplistic and more accessible search phrase query without using search tags for our exploration purposes. Recent systematic or literature reviews that have been published in the last few months utilize more extensive queries and use a range of Boolean and search tags in a wider set of databases. Fleuven et al. performed searches on not only PubMed, but also Embase and Scopus.[51] Peiffer-Smadja et al. used a general search query to identify general infectious diseases on PubMed, Embase, Google Scholar, BioXiv, Acm Digital Library, arXiV, and IEEE.[52] Schinkel et al. performed a review similar to ours by searching only PubMed, but they excluded studies that did not have an AUROC statistic.[53] No search query is perfect; in fact, Salvador-Olivan et al. found that almost 93 percent of search strategies in systematic reviews contained at least one error in their respective search queries.[54]

2.6 CONCLUSION

Overall, we found variation in the domain of predictive analytics tools for septic patients, from feature and population size to choice of method or algorithm. However, it is evident that implementing predictive analytics tools are beneficial in the early detection of sepsis or death related to sepsis. Since most of these studies were retrospective, the translational value in the real-world setting should be further investigated as other variables such as changes in workflow may also have an impact on outcome. Additionally, many solely used one dataset, which is not generalizable across institutions, or even within departments. It will be interesting to see if a predictive analytics tool can be built on top of institutions that have implemented a common data model.

Table 2.1. Overview of unique characteristics and differences between methods

(* indicates prospective)

Author	**Year**	**Goal**	**Population location and size**	**Feature set size**	**"Gold standard" definition**	**Model and performance metrics**
Carrara et al.[8]	2015	Mortality prediction in septic shock patients	MIMIC II (ICU) 30,000+ patients	30 variables	Septic Shock: 1991 SIRS criteria, ICD-9 code for septic shock, abnormal interval must exceed 5hrs for each feature, SIRS 2+, SIRS with low SBP despite adequate fluid resuscitation	Multivariate linear regression with Shrinkage Techniques model Mean square error (MSE): 0.03
Danner et al.[9]	2017	Assess the value of HR-to-systolic ratio in the accuracy of sepsis prediction after ED presentation	Local (ED) 53,313 patients	9 vitals/variables	Sepsis: Discharge diagnosis of sepsis, evaluated vitals, demographics, chief complaints	Multivariate linear regression model - Accuracy: 0.74 - HR to systolic ratio accounted for 69% of overall predictive ability

Capp et al.[10]	2015	Describe key patient characteristics present within 4 hours of ED arrival that are associated with developing septic shock between 4 and 48 hours of ED arrival.	Local (ED) 1,316 patients	5 risk factors	Sepsis: manual chart review with SIRS 2+, evidence of infection (excluded if gastrointestinal bleed) Septic shock: SBP > 90mmHg despite appropriate fluid hydration of 30cc/kg with presence of hypotension for at least 2 hours after	Multivariable logistic regression model Found risk factors associated with progression of sepsis to septic shock between 4 and 48 hours of ED arrival: - Female: 1.59 odds ratio (OR) - Non-Persistent hypotension: 6.24 OR - Lactate > 4 mmol/L: 5.30 OR - Bandemia > 10%: 2.60 OR - Past medical of coronary heart disease: 2.01 OR

Faisal et al.[11]	2018	To develop a logistic regression model to predict the risk of sepsis following emergency admission using the patient's first electronically recorded vital signs and blood test results and to validate this novel computer-aided risk of sepsis model, using data from another hospital.	Local (ED) 57,243 patients	12 vitals/variables	Sepsis: ICD-10 codes without organ failure Severe sepsis: ICD-10 codes with 1+ organ failure or septic shock	Logistic regression models All area under the receiver operator curve (AUROC): 0.79 Sepsis AUROC: 0.70 Severe sepsis AUROC: 0.81
Ho et al.[12]	2012	Investigate how different imputation methods can overcome the handicap of missing information.	MIMIC II (ICU) Sample size not stated	6 vitals	Sepsis: ICD-9 Septic shock: examined clinical chart records	- Sepsis: Multivariate logistic regression models - Septic shock: multivariate logistic regression, linear kernel SVM, and regression trees H: Clinical history feature set P: initial physiological state feature set Sepsis AUROC (imputed mean

						and matrix factorization-based approaches) All H: 0.791 (0.792) Stepwise H: 0.790 (0.791) All H ∪ P: 0.821 (0.822) Stepwise H ∪ P: 0.823 (0.823) Septic shock AUROC: 0.773-0.786
Langeley et al.[13]	2013	Examine clinical features, plasma metabolome, and proteome of patients to predict patient survival of sepsis.	CAPSOD (ED) 1,152 individuals with suspected, community acquired sepsis; Discovery set of 150 patients	4 vitals/variables	Acute infection + 2+ SIRS	Logistic regression (sepsis prediction) and SVM model (survival and death prediction) Logistic regression AUROC: 0.847 Logistic

						regression accuracy: 0.851 *best stats occurred at enrollment SVM AUROC: 0.740 SVM accuracy: 0.746
Sutherland et al.[14] *	2011	Use gene expression biomarkers to prospectively distinguish patients with sepsis from those who experience systemic inflammation from healing of surgery	Local (ICU) 85 patients	42 biomarkers	Likely enter sepsis cohort if met ACCP/SCCM consensus statement and clinical suspicion of systemic infection Confirmation performed retrospectively	Classifier: Recursive partitioning, LASSO, logistic regression. Individual genes examined via Bayes-adjusted linear model. MT-PCR diagnostic classifier generated using a LogitBoost ML algorithm (tree-based) PCR Accuracy: 92% AUROC: 0.86-0.92

Gultepe et al.[15]	2014	Develop a decision support system to identify patients with hyperlactatemia and to predict mortality from sepsis using predicted lactate levels	Local (ED) 741 patients	7 vitals/labs	Sepsis: determined from EHR diagnosis and SIRS criteria	SVM classifier Accuracy: 0.73 AUROC: 0.73
Horng et al.[16]	2017	To demonstrate the incremental benefit of using free text data in addition to vital sign and demographic data to identify patients with suspected infection in the emergency department	Local (ED) 198,833 control 32,103 cases	12 vitals/variables	ED ICD-9-CM code	Linear SVM and free text models Bag of words AUROC: 0.86 Bag of words sensitivity: 0.78 Bag of words specificity: 0.79 Topic model AUROC: 0.85 Topic model sensitivity: 0.80 Topic model specificity: 0.75

Thottakkara et al.[17]	2015	To compare performance of risk prediction models for forecasting postoperative sepsis and acute kidney injury	Local (in-patient) 50,318 patients	285 variables	Forecast post-op sepsis and acute kidney injury AHRQ definition of "post-op sepsis" and organ failure associated with sepsis was identified by ICD-9CM code for acute organ dysfunction	Comparison of models that used logistic regression, generalized additive models (GAM), naive Bayes, SVM Naive Bayes performed the worst in the comparison; GAMs and SVMs had good performance; PCA feature extraction (reduced to 5 features) improved predictive performance for all models Severe sepsis AUROC: 0.76-0.91

Vieira et al.[18]	2013	Proposed a modified binary particle swarm optimization method for feature selection to predict mortality in septic patients	MEDAN (ICU) 382 patients	Model chooses custom number of features (2-7)	MEDAN dataset prelabelled patients for abdominal septic shock	Support vector machine for mortality prediction Modified binary particle swarm optimization (MBPSO): feature selection MBPSO 12 (28) features: No-FS Accuracy: 72.6% (89%) Accuracy: 76.5% (94.4%)
Ghosh et al.[19]	2016	Predict septic shock for ICU patients using non-invasive waveform measurements	MIMIC II (ICU) 1,519 patients	3 vitals/labs	Sepsis: ICD-9 Septic shock: examining clinical chart records	Coupled hidden Markov models (CHMM) with varying gap interval and observation window sizes CHMM average: 0.85 Multi-channel patterns (MCP)-CHMM average: 0.86

Peelen et al.[20]	2009	Develop a set of complex Markov models based on clinical data to extract meaningful clinical patterns and to provide prediction for sepsis and other diseases.	Local (ICU) 2,271 patients	6 variables	Sever sepsis: SIRS 2+ within 24 hrs. of ICU admission and 1+ dysfunctioning organ system (SOFA)	3 Markov models (amount of organ failure, type of organ failure, differences between development and persistence of organ failure) ICU death the error rates were 17.7%, 18.1% and 17.8% and the AUCs were 0.79, 0.79, and 0.80 for Models I, II, and III.
Stanculescu et al.[21]	2014	Demonstrate that by adding a higher-level discrete variable with semantics sepsis/non-sepsis, can detect changes in the physiological factors that signal the presence of sepsis	Local (NICU) 24 neonates	Bradycardia, desaturation	Lab result of blood culture for neonatal sepsis	Hierarchical switching linear dynamical system (HSLDS) Autoregressive (AR)-HMM AUROC: 0.72 HSLDS deep learning AUROC: 0.69 HSLDS known factors AUROC: 0.62

Stanculescu et al.[22]	2014	Detect and identify sepsis in neonates before a blood sample is drawn. Furthermore, they wanted to identify which physiological event would contribute most for detecting sepsis.	Local (NICU) 24 neonates	6 vitals/variables	Positive cultures as pathogens: proven sepsis Positive cultures as mixed growth/skin commensal: "suspected sepsis"	AR-HMM AUROC: 0.74-0.75 AUROC with missing data: 0.72-0.73 AUROC with bradycardia and mini bradycardia: 0.79-0.80 AUROC with desaturation: 0.76-0.78 AUROC with all states: 0.79-0.80
Gultepe et al.[23]	2012	Use a Bayesian network to detect sepsis early	Local (ICU) 1,492 patients	BN1: 5 variables BN2: 7 variables	"Sepsis occurrence"	Bayesian network (BN) models BN-1 (vitals) goodness of fit: 15.4 BN-2 (vitals+MAP) goodness of fit: 19.9 Found that lactate is a driver in both models

						and maybe an important feature for early sepsis detection
Nachimuthu and Haug[24]	2012	Detect sepsis right after patients are admitted to the ED.	Local (ED) 3,100 patients	11 vitals/variables	Clinician determined 'sepsis' during retrospective chart review	Dynamic Bayesian network 3 hrs. after admission AUROC: 0.911 6 hrs. after admission AUROC: 0.915 12 hrs. after admission AUROC: 0.937 24 hrs. after admission AUROC: 0.944
Calvert et al.[25]	2016	Detect and predict the onset of septic shock for alcohol-use disorder patients in the ICU	MIMIC III (ICU) 1,394 patients	9 vitals/variables	Septic shock: SIRS 2+, ICD-9, organ dysfunction, SBP <90 mmHg for 1 hour, total fluid replacement >= 1200mL or	*InSight* Sensitivity: 0.93 Specificity: 0.91 Accuracy: 0.91 F1 Score: 0.161

					20mL/kg for 24 hours	
Calvert et al.[26]	2016	To develop high-performance early sepsis prediction technology for the “general patient population”	MIMIC II (ICU) 29,083 patients	10 vitals/variables	Sepsis: ICD-9 code, 1991 SIRS for 5 hours	*InSight* Sensitivity: 0.90 Specificity: 0.81 AUROC: 0.92 Accuracy: 0.83
Desautels et al.[27]	2016	To validate InSight with the new Sep-3 definition and make predictions using minimal set of variables	MIMIC III (ICU) 22,583 patients	8 vitals/labs	Sepsis: Sep-3 definition, suspicion of infection equated with an order of culture lab draw and dose of antibiotics	*InSight* AUROC: 0.88 APR: 0.60

Mao et al.[28]	2017	Validate the InSight algorithm for detection and prediction of sepsis and septic shock	MIMIC III (ICU) Local (ED, General) 61,532 stays	6 vitals/labs	Sepsis: ICD9 + SIRS 2+ (995.91) Severe sepsis: ICD9 (955.92), organ dysfunction, SIRS 2+ Septic shock: ICD9 (785.52), SBP < 90mmHg (at least 30 min), resuscitated with >= 20 mL/kg over 24 hrs, >=1200mL in total fluids	*InSight* Detect sepsis AUROC: 0.92 Detect severe sepsis AUROC: 0.87 Detect 4 hrs before onset sepsis AUROC: 0.96 Detect 4 hrs. before onset severe sepsis AUROC: 0.85
McCoy et al.[29]	2017	Aimed to improve sepsis-related patient outcomes through a revised sepsis management approach	Local (ICU) 407 patients	6 vitals/variables	Severe sepsis: SIRS 2+, qSOFA score	*Dascena* Sep-3 AUROC: 0.91 Sep-3 sensitivity: 0.83 Sep-3 specificity: 0.96 Severe sepsis AUROC: 0.96 Severe sepsis sensitivity: 0.90 Severe sepsis specificity: 0.85

Shimabukuro et al.[30] *	2017	Randomized control trial to show lowered mortality and length of stay using a machine learning sepsis prediction algorithm	Local (ICU) 75 controls 67 cases	7 vitals/labs	Severe sepsis: 'organ dysfunction caused by sepsis' Random allocation sequence to put patients in groups	*InSight* AUROC: 0.952 Sensitivity: 0.9 Specificity: 0.9 Average length of stay decreased from 13 to 10.3 days In-hospital mortality decreased by 12.3%
Henry et al.[32]	2015	Create and test a score that predicts which patients will develop septic shock	MIMIC II (ICU) 16,234 patients	54 features	Suspicion of infection: ICD-9 or by presence of clinical note that mentioned sepsis or septic shock Sepsis: suspicion + SIRS Severe sepsis: sepsis + organ dysfunction	TREWScore (Cox proportional hazards model using the time until the onset of septic shock as the supervisory signal) AUROC: 0.83 Specificity: 0.67 Sensitivity: 0.85 "Patients were identified a median of 28.2 hours before shock onset"

Lukaszewski et al.[31] *	2008	Detect and identify septic patients before displaying symptoms for ICU patients	Local (ICU) 92 patients	7 cytokines	Admission diagnosis upon ICU entry	Neural networks using cytokine and chemokine data Sensitivity: 0.91 Specificity: 0.80 Accuracy: 0.95
Nemati et al.[35]	2017	Aimed to develop and validate an artificial intelligence sepsis algorithm for early prediction of sepsis	Local (ICU) 33,069 patients	65 variables	Sepsis: Sepsis-3	Modified Weibull-Cox proportional hazards model 4 hrs. in advance AUROC: 0.85
Pereira et al.[33]	2011	Examined different approaches to predicting septic shock with missing data.	MEDAN (ICU) 139 patients	2 sets of 12 and 28 "selected features"	Septic shock: associated with abdominal causes (not clearly defined, data may be prelabelled)	Zero-Order-Hold (ZOH) Fuzzy c-means clustering based on partial distance calculation strategy (FCM-PDS) Performance improvements occur where up to 60% of the data is missing ZOH-FCM-PDS 12 (28) feature AUROC: 0.899

						(0.649); FCM-PDS 12 (28) feature AUROC: 0.786 (0.631)
Ribas et al.[36]	2011	Demonstrate that a SVM variant can provide automatic ranking of mortality predictor and have higher accuracy that current methods.	Local (ICU) 354 patients	4 vitals/labs	Severe sepsis: organ dysfunction (SOFA)	Relevance vector machine AUROC: 0.80 Error rate: 0.24 Sensitivity: 0.66 Specificity: 0.80
Sawyer et al.[38] *	2011	Evaluate if implementing an automated sepsis screening and alert system can facilitate in early interventions by identifying non-ICU patients at risk for developing sepsis	Local (Non-ICU) 270 patients	9 vitals/variables	Intervention group: real-time sepsis alert generated from Clinical Desktop	Recursive partitioning regression tree analysis Within 12hrs of sepsis alert, 70.8% of patients in the intervention group received treatment vs. 55.8% in control.

Shashikumar et al.[37]	2017	Investigates the utility of high-resolution blood pressure and heart rate times series dynamics for the early prediction of sepsis	Local (ICU) 242 patients	11 vitals/variables	Sepsis: Seymour (Sep-3) at some point during ICU stay	Elastic Net logistic classifier: 3 models: (1) entropy features, (2) EMR + socio-demographic-patient history features, (3) models 1+2 Model 1 AUROC (Accuracy): 0.67 (0.47) Model 2 AUROC (Acc): 0.70 (0.50) Model 3 AUROC (Acc): 0.78 (0.61)
Taylor et al.[39]	2016	Compare a machine learning approach to existing clinical decision rules to predict sepsis in-hospital mortality	Local (ED) 4,676 patients	20 variables	ICD-9 with AHRQ clinical classification software to obtain more exhaustive list of patients	Random forest model AUROC: 0.86

Table 2.2. Count of Methods Used Amongst Analyzed Studies

Method	**Retrospective Count**	**Prospective Count**
InSight/Dascena	5	1
Regression models	6 (4 logistic, 2 linear)	1 (logistic)
Markov models	4 (3 HMM, 1 MM)	-
Vector machine	4 (3 SVM, 1 RVM)	-
Bayesian network	2	-
Hazard models	2	-
Neural network	-	1
Fuzzy c-means clustering	1	-
Regression tree	-	1
Net classifier	1	-
Linear dynamic	1	-
Random forest	1	-
Total	13	

2.7 REFERENCES FOR CHAPTER 2

1. Data & Reports | Sepsis | CDC. Centers for Disease Control and Prevention. https://www.cdc.gov/sepsis/datareports/index.html. Accessed August 11, 2019.

2. O'Brien J. The Cost of Sepsis. Centers for Disease Control and Prevention. https://blogs.cdc.gov/safehealthcare/the-cost-of-sepsis/. Published November 18, 2016. Accessed August 11, 2019.

3. Sepsis. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/sepsis/symptoms-causes/syc-20351214. Published November 16, 2018. Accessed August 11, 2019.

4. Balk RA. Systemic inflammatory response syndrome (SIRS): Where did it come from and is it still relevant today? Virulence. 2014;5(1):20-26. doi:10.4161/viru.27135

5. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016;315(8):801. doi:10.1001/jama.2016.0287.

6. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. BioMedical Engineering OnLine. 2014;13(1):94. doi:10.1186/1475-925X-13-94

7. Tsang-Hsiang Cheng, Chih-Ping Wei, Tseng VS. Feature Selection for Medical Data Mining: Comparisons of Expert Judgment and Automatic Approaches. In: 19th IEEE Symposium on Computer-Based Medical Systems (CBMS'06). Salt Lake City, UT: IEEE; 2006:165-170. doi:10.1109/CBMS.2006.87

8. Carrara M, Baselli G, Ferrario M. Mortality prediction in septic shock patients: Towards new personalized models in critical care. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Milan: IEEE; 2015:2792-2795. doi:10.1109/EMBC.2015.7318971

9. Danner OK, Hendren S, Santiago E, Nye B, Abraham P. Physiologically-based, predictive analytics using the heart-rate-to-Systolic-Ratio significantly improves the timeliness and accuracy of sepsis prediction compared to SIRS. The American Journal of Surgery. 2017;213(4):617-621. doi:10.1016/j.amjsurg.2017.01.006

10. Capp R, Horton CL, Takhar SS, et al. Predictors of patients who present to the emergency department with sepsis and progress to septic shock between 4 and 48 hours of emergency department arrival. Critical Care Medicine. 2015;43(5):983–988. doi:10.1097/CCM.0000000000000861

11. Faisal M, Scally A, Richardson D, et al. Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test

Results*: Critical Care Medicine. 2018;46(4):612-618. doi:10.1097/CCM.0000000000002967

12. Ho JC, Lee CH, Ghosh J. Imputation-Enhanced Prediction of Septic Shock in ICU Patients Categories and Subject Descriptors. HI-KDD 2012: ACM SIGKDD Workshop on Health Informatics. 2012.

13. Langley RJ, Tsalik EL, Van Velkinburgh JC, et al. Sepsis: An integrated clinico-metabolomic model improves prediction of death in sepsis. Science Translational Medicine. 2013. doi:10.1126/scitranslmed.3005893

14. Sutherland A, Thomas M, Brandon RA, et al. Development and Validation of a Novel Molecular Biomarker Diagnostic Test for the Early Detection of Sepsis.; 2011:R149. doi:10.1186/cc10274

15. Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: A machine learning basis for a clinical decision support system. doi:10.1136/amiajnl-2013-001815

16. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. Groza T, ed. PLOS ONE. 2017;12(4):e0174708. doi:10.1371/journal.pone.0174708

17. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, et al. Application of machine learning techniques to high-dimensional clinical data to forecast postoperative complications. PLoS ONE. 2016. doi:10.1371/journal.pone.0155705

18. Vieira SM, Mendonça LF, Farinha GJ, Sousa JMC. Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients. Applied Soft Computing Journal. 2013;13(8):3494–3504. doi:10.1016/j.asoc.2013.03.021

19. 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. 2017. doi:10.1016/j.jbi.2016.12.010

20. Peelen L, de Keizer NF, de Jonge E, Bosman RJ, Abu-Hanna A, Peek N. Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failpure in patients in the Intensive Care Unit. Journal of Biomedical Informatics. 2010;43(2):273–286. doi:10.1016/j.jbi.2009.10.002

21. Stanculescu I, Williams CKI, Freer Y. A Hierarchical Switching Linear Dynamical System Applied to the Detection of Sepsis in Neonatal Condition Monitoring. UAI’14 Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence. 2014:752–761.

22. Stanculescu I, Williams CKI, Freer Y. Autoregressive Hidden Markov Models for the Early Detection of Neonatal Sepsis. IEEE Journal of Biomedical and Health Informatics. 2014;18(5):1560–1570. doi:10.1109/JBHI.2013.2294692

23. Gultepe E, Nguyen H, Albertson T, Tagkopoulos I. A Bayesian network for early diagnosis of sepsis patients: A basis for a clinical decision support system. 2012 IEEE 2nd International Conference on Computational Advances in Bio and Medical Sciences, ICCABS 2012. 2012:1–5. doi:10.1109/ICCABS.2012.6182636

24. Nachimuthu SK, Haug PJ. Early detection of sepsis in the emergency department using Dynamic Bayesian Networks. AMIA . Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2012;2012:653–62.

25. Calvert J, Desautels T, Chettipally U, et al. High-performance detection and early prediction of septic shock for alcohol-use disorder patients. Annals of Medicine and Surgery. 2016;8:50-55. doi:10.1016/j.amsu.2016.04.023

26. Calvert JS, Price DA, Chettipally UK, et al. A computational approach to early sepsis detection. Computers in Biology and Medicine. 2016;74:69–73. doi:10.1016/j.compbiomed.2016.05.003

27. Desautels T, Calvert J, Hoffman J, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. JMIR Medical Informatics. 2016. doi:10.2196/medinform.5909

28. Mao Q, Jay M, Hoffman JL, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open. 2018. doi:10.1136/bmjopen-2017-017833

29. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Quality. 2017. doi:10.1136/bmjoq-2017-000158

30. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, 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. 2017. doi:10.1136/bmjresp-2017-000234

31. Lukaszewski RA, Yates AM, Jackson MC, et al. Presymptomatic prediction of sepsis in intensive care unit patients. Clinical and Vaccine Immunology. 2008;15(7):1089–1094. doi:10.1128/CVI.00486-07

32. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Science Translational Medicine. 2015;7(299). doi:10.1126/scitranslmed.aab3719

33. Pereira RDMA, Fialho AS, Cismondi F, et al. Predicting septic shock outcomes in a database with missing data using fuzzy modeling: Influence of pre-processing techniques on real-world data-based classification. IEEE International Conference on Fuzzy Systems. 2011:2507–2512. doi:10.1109/FUZZY.2011.6007606

34. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Scientific Data. 2016;3(1). doi:10.1038/sdata.2016.35.

35. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Critical Care Medicine. 2017. doi:10.1097/CCM.0000000000002936

36. Ribas VJ, Lopez JC, Ruiz-Sanmartin A, et al. Severe sepsis mortality prediction with relevance vector machines. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Boston, MA: IEEE; 2011:100-103. doi:10.1109/IEMBS.2011.6089906

37. Shashikumar SP, Stanley MD, Sadiq I, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. Journal of Electrocardiology. 2017;50(6):739-743. doi:10.1016/j.jelectrocard.2017.08.013

38. Sawyer AM, Deal EN, Labelle AJ, et al. Implementation of a real-time computerized sepsis alert in nonintensive care unit patients. Critical Care Medicine. 2011;39(3):469–473. doi: 10.1097/CCM.0b013e318205df85

39. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. Jones A, ed. Academic Emergency Medicine. 2016;23(3):269-278. doi:10.1111/acem.12876

40. Clark M. Retrospective versus prospective cohort study designs for evaluating treatment of pressure ulcers: a comparison of 2 studies. Journal of Wound Ostomy & Continence Nursing. 2008;35(4):391-394. doi:10.1097/01.WON.0000326657.57939.c9

41. Raith EP, Udy AA, Bailey M, et al. Prognostic Accuracy of the SOFA Score, SIRS Criteria, and qSOFA Score for In-Hospital Mortality Among Adults With Suspected Infection Admitted to the Intensive Care Unit. JAMA. 2017;317(3):290. doi:10.1001/jama.2016.20328

42. Govindan S, Prescott HC, Chopra V, Iwashyna TJ. Sample size implications of mortality definitions in sepsis: a retrospective cohort study. Trials. 2018;19(1). doi:10.1186/s13063-018-2570-2

43. UCSF Case Study: Dascena. https://www.dascena.com/ucsf. Accessed August 11, 2019.

44. Nassir Ghaemi S, Shirzadi AA, Filkowski M. Publication bias and the pharmaceutical industry: the case of lamotrigine in bipolar disorder. Medscape J Med. 2008;10(9):211.

45. Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nature Methods. 2016;13(9):703-704. doi:10.1038/nmeth.3968

46. Worster A, Fan J, Ismaila A. Understanding linear and logistic regression analyses. CJEM. 2007;9(02):111-113. doi:10.1017/S1481803500014883

47. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Briefings in Bioinformatics. 2018;19(6):1236-1246. doi:10.1093/bib/bbx044

48. Wang H, Wu Z, Xing EP. Removing Confounding Factors Associated Weights in Deep Neural Networks Improves the Prediction Accuracy for Healthcare Applications. 2019:17.

49. Bishop CM. Bayesian PCA. In: Kearns MJ, Solla SA, Cohn DA, eds. Advances in Neural Information Processing Systems 11. MIT Press; 1999:382–388. http://papers.nips.cc/paper/1549-bayesian-pca.pdf.

50. Ravindra K, Rattan P, Mor S, Aggarwal AN. Generalized additive models: Building evidence of air pollution, climate change and human health. Environment International. 2019;132:104987. doi:10.1016/j.envint.2019.104987

51. Schinkel M, Paranjape K, Nannan Panday RS, Skyttberg N, Nanayakkara PWB. Clinical applications of artificial intelligence in sepsis: A narrative review. Computers in Biology and Medicine. 2019;115:103488. doi:10.1016/j.compbiomed.2019.103488

52. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Medicine. January 2020. doi:10.1007/s00134-019-05872-y

53. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clinical Microbiology and Infection. September 2019. doi:10.1016/j.cmi.2019.09.009

54. Salvador-Oliván JA, Marco-Cuenca G, Arquero-Avilés R. Errors in search strategies used in systematic reviews and their effects on information retrieval. Journal of the Medical Library Association. 2019;107(2). doi:10.5195/jmla.2019.567

Chapter 3. A SIMPLIFIED FRAMEWORK TO EXTRACT SOCIAL AND BEHAVIORAL DETERMINANTS OF HEALTH: A DATA SCIENCE APPROACH

3.1 ABSTRACT

Background: Social and behavioral determinants of health can be important indicators of health outcomes. However, collection of these data is not standard and can be commonly found in both structured and unstructured forms, with past literature determining structured data to be less complete.

Objective: To extract, classify, and score a subset of social and behavioral factors from clinical notes using common text classification methods.

Setting: We collaborated with a local Level I trauma hospital located in an underserved area that has a housing unstable patient population of about 6.5% and extracted text notes related to various social and behavioral determinants of health for acute care patients.

Methods: We solely utilized open source Python packages to test simple text classification methods that can potentially be easily generalizable and implemented. We extracted social history text from various sources, such as admission and emergency department notes, over a five-year timeframe and performed manual chart reviews to ensure data quality. We manually labelled the sentiment of the notes, treating each text entry independently. Four different models with two different feature selection methods (bag of words (BOW) and bigrams) were used to classify and predict housing stability, tobacco use, and alcohol use status for the extracted clinical text.

Results: From our analysis, we found overall positive results and metrics in applying open-source classification techniques; the accuracy scores were 91.2%, 84.7%, 82.8% for housing stability, tobacco use, and alcohol use respectively. There were many limitations in our analysis including social factors not present due to patient condition, multiple copy-forward entries and shorthand. Additionally, it was difficult to translate usage degrees for tobacco and alcohol use. However, when compared to structured data sources, our classification approach on unstructured notes yielded more results for housing and alcohol use; tobacco use proved less fruitful for unstructured notes.

Keywords: Text classification, Sentiment modelling, Social factors, Social history, Risk scores

3.2 INTRODUCTION

The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 established guidelines to help improve patient safety and efficacy by laying the framework for electronic health record (EHR) adoption in the United States through financial incentives.[1] With the HITECH Act and incentives through Meaningful Use, EHR adoption skyrocketed and large databases of clinical information were implemented.[2] These large databases can contain simple information such as patient demographics and vital signs, but it can also contain more qualitative or descriptive data such as clinical notes and images. With Meaningful Use, the completeness of the data being collected has increased. Currently, many institutions have large amounts of underutilized data that are ideal for biomedical exploration and discovery to aid in patient care, creating new exploration opportunities.

Most data can be generally categorized as structured or unstructured, where structured data can consist of items such as vital signs and lab results and unstructured data can consist of items such as text notes, images, or multimedia.[3] Structured data have been essential in modern databases as they are significantly easier to query, merge, or filter when sifting through the data. They have two parts which simplifies the search process: (1) variable name and (2) value.[4]

Structured data can be easily added and expanded and has proven critical in modern clinical databases, especially for data such as patient vitals and demographics. On the other hand, although structured data can generally be easier to extract and analyze, unstructured data can potentially provide an array of information not present or easily identifiable in structured data. Challenges arise with unstructured data as they are not as easily interpretable as or categorizable as a numeric structured value. Images and text often contain many levels of metadata that would need manual review or more advanced extraction techniques to decode or interpret; basic queries will not work in these cases as the data are simply not queryable. Furthermore, if a provider is uncertain, CMS guidelines state that they "do not code diagnosis documented probable, suspected, questionable, rule out, compatible with, consistent with, or working diagnosis or similar terms indicating uncertainty"; however, a statement could be written in the provider notes for reference.[5] This type of detailed encounter information would not be available in structured data formats. Additionally, clinicians have recently expanded intake data and social determinants of health (SDoH) information are starting to become more readily available. Furthermore, there has a been a growing interest around Medicaid patients, as SDoH can drive up to 80% of health outcomes, especially within this patient demographic.[6] Therefore, SDoH and REAL (Race, Ethnicity and Language) data are now starting to be analyzed for secondary research as recent research has indicated that there is a correlation between SDoH and health outcomes and the increasing need to research health disparities across populations.[7]

SDoH and REAL can include housing stability, access jobs and health care services, education level, language, and socioeconomic conditions.[8] These indicators are population descriptors which are useful as health outcomes predictors and the utilization of health interventions.[9] Because they can potentially be strong indicators of health and health outcomes, many are now starting to analyze and increase intake of SDoH and REAL information, whether through text notes or standardized coding, such as International Classification of Diseases (ICD).[10,11] However, there are challenges with SDoH intake as there is no standardized SDoH screening tool in the EHR[12]. Furthermore, many times these data are self-reported or inferred by the provider, especially if the patient is in a noncommunicable state, such as experiencing a loss of consciousness. Additionally, coding schemes, like ICD, can be unreliable in secondary analysis as coding can oversimplify symptoms and diagnoses into general coding categories, leading to uncertainties and coding errors may be present from unintentional mistakes or even upcoding.[13,14] Past research has shown that hospital readmissions are highly influenced by patient health status and SDoH and suggest that clinical staff and researchers should consider SDoH when assessing readmission risk.[15]

The 2018-2019 King County Community Health Needs Assessment (CHNA) reported the results from a health needs assessment survey given to residents to identify regional perceived healthcare issues. It was determined that housing affordability and housing stability were major challenges dominating overall health. Mental health was also highlighted as a challenge for healthcare providers; mental illness can be caused by depression, schizophrenia, and alcohol and substance-related disorders. The CHNA reported that adults in the lowest income tier were about 15 times more likely to experience severe psychological distress compared to their high-income counterparts. Additionally, it noted that part of the region had continued challenges with adult smoking rates.[16]

Locally, there were more than 12,000 people experiencing housing instability in 2018.[17] Because housing instability is often associated with other health inequalities as “the social determinants of homelessness and health inequities are often intertwined, and long term homelessness further exacerbates poor health”,[18] it is therefore important to treat housing stability and other SDoH as a combined health issue to aid in improving health outcomes in clinical settings. Additionally, various behavioral health habits, including tobacco and alcohol use, although may not directly be considered a SDoH, can impact health decisions and outcomes. For example, participants from a study who drank alcohol and self-reported tobacco use tended to have a diet higher in fat and sugar, low in vitamins and minerals as well as less healthy.[19] Within our region, it has been noted in recent years that the smoking rate is around 13 percent; however, among Black/African-Americans or individuals with multiple races, is double the rate among white adults and four times higher than Asian adults. Additionally, it was reported that, when compared to high income households, low income households were three times more likely to be smokers.[16,20] Drug and alcohol use also shared similar metrics; within the region, “drug and alcohol-caused deaths was 22% higher among Blacks and four times greater among American Indian/Alaskan Native than among non-Hispanic Whites” and alcohol use represented 4.97 per 100,000 deaths locally in 2015.[21,22] Therefore it may be important to look at the combined category of social and behavioral determinants of health (SBDH) to better understand the patient population.[19]

Recent technological advances in machine learning and artificial intelligence have shown great potential in providing a pathway for informaticians and clinicians to better understand unstructured data. Within the clinical setting, there have been numerous approaches in adopting natural language processing (NLP) to aid with processing unstructured clinical text notes. Common uses of NLP include extracting diagnoses and chief complaints as well as grouping of information for quality improvement. There are various NLP methods that can be used in the clinical setting, such as automatic tagging of conditions or variables of interest, sentiment classification, or even text extraction. Various open source NLP and ontological tools, such as Automated Retrieval Console, Apache clinical Text Analysis and Knowledge Extraction System (Apache cTAKES), MetaMap, and HITEx, Unified Medical Language System (UMLS) Metathesaurus and BioPortal have been used to aid with text extraction or classification.[23–25] On the other hand, less complex classification methods have been used as well to identify specific groups of patients, risk assessment, or aid in validating structured annotation.[26,27,28] A recent scoping review found that although practitioners collect a variety of SBDH data at point of care through EHR, the overall use of automated technology is limited to date.[29]

With the idea of implementing an easily generalizable approach to classify selected social factors, we extracted both unstructured and structured data sources related to SBDH from a local hospital to identify and generate a framework to automatically extract and classify SBDH from text notes. We focused on housing stability status, tobacco use, and alcohol use. These three social factors were chosen due to their direct impact on health outcomes and the local public health impact[17–19,30,31] and presence in the EHR. To tackle challenges associated with SBDH extraction from unstructured text notes, we aimed to create a generalizable framework using low barrier open-source tools that are commonly used in the data science field. Because notes and stylistic choices can be institution and location specific, we sought not to create a model that is generalizable but rather a simplified method that could be potentially easily implemented using common off the shelf NLP and data science tools.

3.3 METHODS

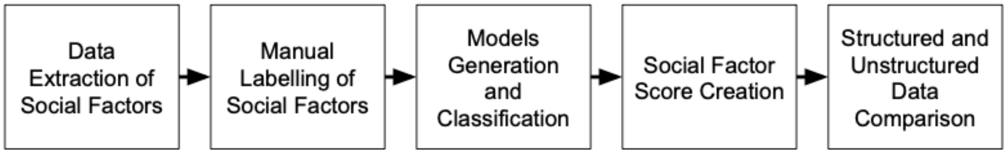


Figure 3.1. Workflow diagram.

3.3.1 *Study Design and Overview*

A high-level overview of our workflow can be seen in Figure 3.1. We conducted a retrospective cohort study of patients in the acute care setting at a Level I trauma center and academic teaching hospital with the aim to create a general and easily applicable workflow to extract and classify social factors from clinical notes. We applied a two-pronged approach and collected unstructured data from a subset of patients over a 1-year timespan (Group A) to create and test the text classification model and also collected structured and unstructured data from a subset of patients over a 5-year timespan (Group B) to apply the best model created from Group A and compare results between the two data types. We performed automatic classification and scoring of patients via various NLP classification methods on three social factors: (1) housing stability, (2) tobacco use, and (3) alcohol use. Our general workflow for housing stability, a similar approach was also used for tobacco and alcohol use, can be seen in Figure 3.2. Patient data were extracted directly from the data warehouse and stored on encrypted computers and were not distributed or shared outside of the secured and closed environment.

3.3.2 *Study Population*

Data were extracted from Harborview Medical Center, a 413-bed academic hospital that has a patient population consisting mostly from Washington, but also from a five-state area.[32] In 2014, there were 17,121 inpatient admissions, where 19 percent of the patients belong to a racial or ethnic minority and 37 percent of patients were enrolled in Medicaid.[32,33] Additionally, in 2015, the non-US born population was estimated to be around 21 percent in Seattle, highlighting the potential diversity that could be found with this patient population.[33]

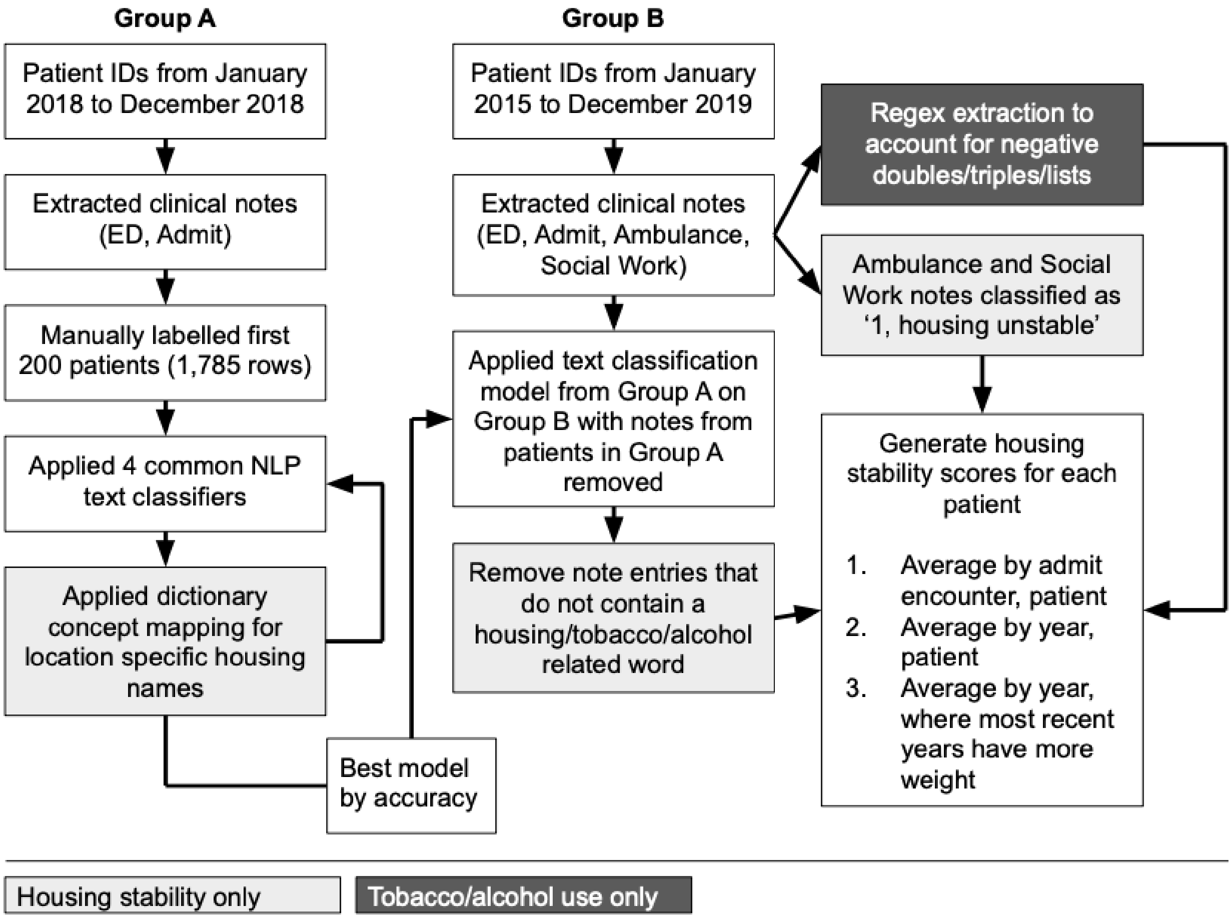


Figure 3.2. Housing stability extraction diagram.

3.3.3 *Data Sources, Extraction, and Validation*

We extracted both structured and unstructured data sources related to housing stability, tobacco use, and alcohol use. These SBDH indicators were selected as they were important indicators highlighted in the King County CHNA and because they were more readily available in the clinical notes. Extraction was performed using SQL queries called directly from an integrated python-based Jupyter Notebook:

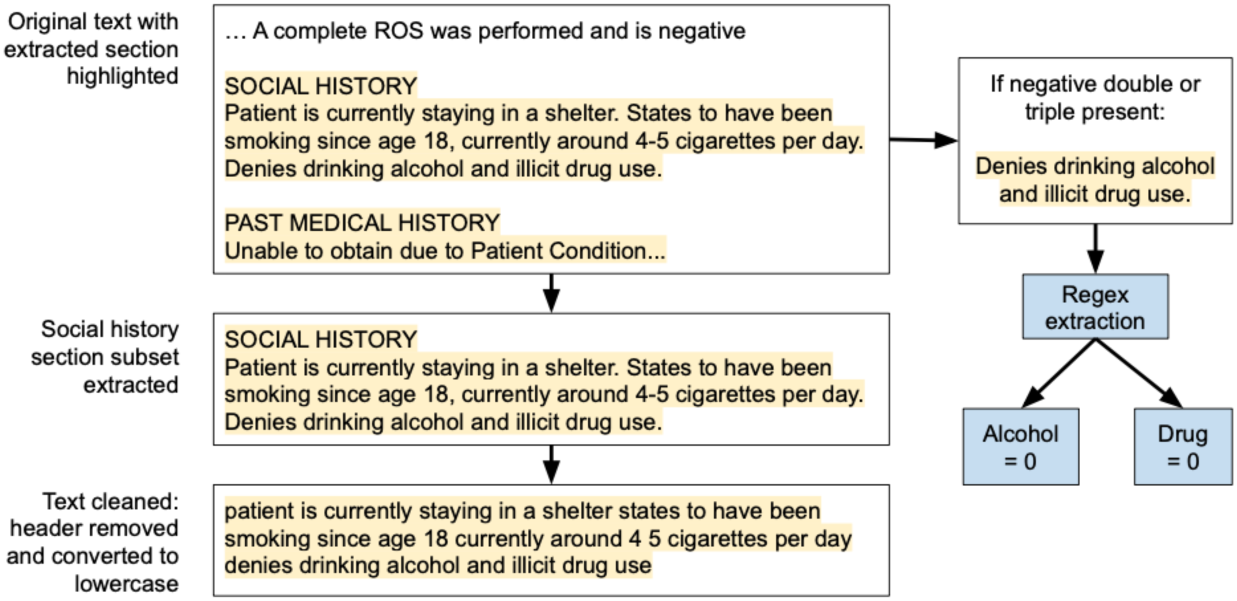
a. Structured data sources include billing and diagnostic/International Classification of Disease (ICD) 9 and 10 codes, questionnaire or Epic SmartForm responses, address fields (location), problem list (ICD 9), patient encounters, clinical events (actual encounters of care), and discharge/disposition location.
b. Unstructured data sources consisted of text notes from the emergency department (ED), admission (admit) notes, social work, and ambulance notes.

Discharge notes were not explored as they were not recorded in the same subdivided format as the admit and ED notes, making selective text extraction of SBDH difficult. From our initial list of patient identifiers over a one-year timespan from Group A, we performed manual EHR validation of a random subset of 25 patients to validate the completeness of the clinical notes and confirm the location of social history and social factors in clinical notes. Extensive research and conversations with an internal data analyst confirmed the location of these topics (housing, tobacco use, and alcohol use) within structured data sources.

3.3.4 *Data Cleaning*

After confirmation, clinical notes were extracted for both Groups A and B. The notes were cleaned (e.g. symbols removed, converted to lowercase) prior to classification and analysis in the Jupyter notebook via NLTK. Our general text extraction and cleaning workflow can be seen in Figure 3.3. However, housing stability notes and tobacco or alcohol use notes were stylistically and grammatically different, and both sets needed distinct additional cleaning steps. Housing stability notes that contained the phrase ‘not homeless’ were converted via regex to say ‘housed’ instead. Additionally, for housing stability, a concept dictionary was also created to substitute local facility names with more general concept (e.g. ‘Union Gospel Mission’ was converted to ‘shelter’). This was done to explore how the algorithms handle formal nouns.

For text notes in Group B, we performed an additional concept extraction step. Tobacco and alcohol use notes often contained incomplete (lacking the subject, predicate, object format) triples or doubles (e.g. ‘Denies smoking, drinking, drugs’). Due to their incomplete sentence structures, common NLP tools to parse, extract, and classify triples, such as Stanford CoreNLP, were not suitable as these tools rely on having all three parts of the triple present. These notes related to tobacco and alcohol use therefore underwent an additional step that performed a separate relation extraction that first pulls out the SBDH related objects and then would reclassify and label the negative sentiment to all components of the list. Our process can be seen in the left side of Figure 3.3. If the regex extraction of negative lists resulted in a different result from the text classification prediction, the regex extraction would overwrite the end result prior to scoring. Once these steps were performed, the data were considered clean and suitable for classification.



… A complete ROS was performed and is negative

SOCIAL HISTORY
Patient is currently staying in a shelter. States to have been smoking since age 18, currently around 4-5 cigarettes per day. Denies drinking alcohol and illicit drug use.

PAST MEDICAL HISTORY
Unable to obtain due to Patient Condition…

SOCIAL HISTORY
Patient is currently staying in a shelter. States to have been smoking since age 18, currently around 4-5 cigarettes per day. Denies drinking alcohol and illicit drug use.

patient is currently staying in a shelter states to have been smoking since age 18 currently around 4 5 cigarettes per day denies drinking alcohol and illicit drug use

Denies drinking alcohol and illicit drug use.

Figure 3.3. Text cleaning workflow.

3.3.5 *Model Building*

Cleaned text from Group A were used to generate and test the classification models. These notes were split in 70/30 validation and testing sets. We applied four different common NLP text classification models to the testing sets (via SciKit Learn): multinomial naïve Bayes, support vector machine, logistic regression, and random forest. Default parameters and a bag-of-words approach were used. The best performing model by accuracy was then chosen and applied to the larger corpus, Group B, with notes from patients in Group A removed, to avoid overfitting and classification bias. This process was performed for housing, tobacco use, and alcohol use.

3.3.6 *Scoring Generation*

In order to create a simple method of identifying patients who are experiencing social instability, we created a scoring metric based on the classified notes. After applying the optimum model by accuracy to the entire corpus of extracted text notes, housing stability, tobacco use, and alcohol use scores were generated. Patient identifiers were mapped by patient location and those who were not in the acute care setting during this timeframe were removed. Three different scoring approaches were used to describe these social factors: (1) predictions were averaged by patient encounter, then averaged by patient identifier, (2) predictions were averaged by year, then by patient identifier, and (3) predictions were averaged by year, where each year then had a weight where the most recent year had the highest weight and the furthest year had the lowest weight (e.g. predictions from 2019 were weighted by a factor of 5 and predictions from 2015 were weighted by a factor of 1). This scoring generation process was then repeated on our structured data for all three social factors and the results were compared and analyzed. Structured data was also extracted for our list of patients in Group B.

3.4 RESULTS

3.4.1 *Characteristics of Study Participants*

Clinical notes (ED, admit, social work, and ambulance) between 2015 and 2019 were extracted and included, forming Group B. Notes from the first 200 patients were included in Group A and notes from 147,457 patients were included in Group B. During the same timeframe, 61,767 patients were in acute care. After extraction and model prediction, the patient notes were cross referenced with inpatient location and only notes from those who were in acute care were retained, for a total of 43,798 patients from 2015 to 2019. The patient demographics of this final subset were 63% (*n*=27,575) male, 37% (*n*=16,223) female, 88.2% (*n*=38,634) not Hispanic or Latino, and 10.5% (*n*=4,609) Hispanic or Latino, and 1.3% (*n*=555) unknown or not answered. Further descriptive statistics can be found in Table 3.1.

Table 3.1. Population demographics

Race (*n*=43,798)	*n* (%)
White or Caucasian	31,575 (72.1%)
Black or African American	4,812 (11.0%)
Asian	3,174 (7.2%)
American Indian or Alaska Native	1,165 (2.7%)
Native Hawaiian or other Pacific Islander	524 (1.2%)
Multiple races	3 (0%)
Unavailable, unknown, or missing	2,545 (5.8%)
Age range (*n*=43,798)	*n* (%)
0-18	1,856 (4.2%)
19-44	12,437 (28.4%)
45-64	14,863 (33.9%)
65-84	11,902 (27.2%)
85 and over	2,740 (6.3%)

3.4.2 *Data Attributes*

Table 3.2 illustrates the amount of data for each corresponding extraction level, specifically for housing status. We first started with extracting text from the ED and admit notes, forming Group A, which consisted of 50,000 rows or text entries and covered 3,200 unique patients, over a one-year timeframe. From there, we manually labelled housing stability concepts in a binary fashion, where 0 would indicate housing stability and 1 would indicate any level of housing instability, regardless of severity. As manual labelling can be a labor-intensive process, only the first 6,000 text rows were labelled, covering 218 unique patients. However, within these first 6,000 rows, numerous notes did not contain text that alluded to housing status or were empty due to patient condition. Therefore, only 1,785 out of the 6,000 rows were labelled, covering 200 unique patients, where 995 (55.7%) were labelled as housing stable and 790 (44.3%) were labelled as housing unstable. We also found that 5.7% of the entries within this subset were duplicates or copy-forward entries. The same workflow was performed for labelling tobacco and alcohol use. However, only 1,108 rows were labelled for tobacco use and 1,220 rows for alcohol use, where in both cases 0 indicated no use, 1 indicated rare/previous/occasional use, and 2 indicated current use, regardless of degree. Tobacco use resulted in 446 (40.3%) labels for no use, 129 (11.6%) labels for rare/previous/occasional use, and 533 (48.1%) labels for current use. Similarly, alcohol use resulted in 595 (48.8%) labels for no use, 185 (15.2%) labels for rare/previous/occasional use, and 440 (36%) labels for current use.

Table 3.2. Extracted data amounts for housing status

Level of extraction	Rows (*n*)	Unique patients (*n*)	Unique encounters (*n*)	Social history entries (*n*/unique)
ED and Admit notes	49,955	3,233	15,664	21,876/21,334
Housing, Tobacco, Alcohol Information	6,000	218	1,995	2,408/2,211

Remove nulls/missing data	Housing: 1,785 Tobacco: 1,108 Alcohol: 1,220	Housing: 200 Tobacco: 179 Alcohol: 181	1,361	1,785/1,684

3.4.3 *Model Performance*

Four different common text classifiers, mentioned in the Methods section, were applied to the manually labelled Group A data. The statistical metrics, including accuracy, precision, and recall, can be seen in Table 3.3 and 3.4. The accuracies between the classifiers and each classification technique for housing stability were overall fairly high ranging from 84.4-92.2%. The accuracies for tobacco and alcohol use were lower, ranging from 70.9-84.7% for tobacco use and 70-82.8% for alcohol use. Additionally, for each top performing model, the most influential words for text classification, for each social factor, can be seen in Table 3.5. The best performing classification models were selected for each social factor and were used to apply the model to our entire corpus in Group B.

Table 3.3. Accuracies amongst text classifiers

	n=1	n=1-2
Multinomial naïve Bayes	Housing: 91.62% Tobacco: 70.87% Alcohol: 70.77%	Housing: 91.43% Tobacco: 77.18% Alcohol: 69.95%
Support vector machine	Housing: **92.18%** Tobacco: 81.08% Alcohol: 76.50%	Housing: 91.99% Tobacco: 82.88% Alcohol: 81.97%
Logistic regression	Housing: 84.36% Tobacco: 75.38% Alcohol: 77.60%	Housing: 90.13% Tobacco: **84.68%** Alcohol: **82.79%**
Random forest	Housing: 90.50% Tobacco: 76.28% Alcohol: 71.31%	Housing: 91.25% Tobacco: 78.98% Alcohol: 75.68%

Table 3.4. Best performing classifier detailed metrics

	Classifier	Accuracy	Recall	Precision	F1
Housing status*	Support vector machine (n=1)	0.92	0.93/0.91 (0/1)	0.94/0.90	0.93/0.91

Tobacco use**	Logistic Regression (n=1-2)	0.85	0.82/0.95/0.86 (0,1,2)	0.96/0.43/0.87 (0,1,2)	0.88/0.60/0.87 (0,1,2)
Alcohol use**	Logistic Regression (n=1-2)	0.83	0.86/0.73/0.81 (0,1,2)	0.93/0.44/0.88 (0,1,2)	0.89/0.55/0.84 (0,1,2)

* 0: no use, 1: current use

** 0: no use, 1: rare/occasional/history, 2: current use

Table 3.5. Importance ranking

Social factor (Classifier)	Top 20 weighted words
Housing stability (support vector machine, n=1)	['friends' 'motel' 'stay' 'cigs' 'found' 'street' 'stays' 'streets' 'van' 'incarcerated' 'desc' 'currently' 'undomiciled' 'friend' 'respite' 'kcj' 'shelters' 'homelessness' 'shelter' 'homeless']
No tobacco use (logistic regression, n=1,2)	['use denies' 'deneis' 'lives' 'tobacco drug' 'seattle denies' 'use results' 'lives seattle' 'alcohol tobacco' 'tobacco drugs' 'never smoker' 'etoh tobacco' 'drinking' 'seattle tobacco' 'denies cigarettes' 'drugs tobacco' 'denies alcohol' 'tobacco alcohol' 'denies smoking' 'denies' 'denies tobacco']
No alcohol use (logistic regression, n=1,2)	['care' 'ppd' 'tobacco' 'smoking' 'etoh tobacco' 'history cocaine' 'tobacco alcohol' 'etoh illicit' 'alcohol tobacco' 'etoh drug' 'drugs etoh' 'alcohol drug' 'use none' 'alcohol drugs' 'drug etoh' 'denies alcohol' 'lives' 'denies drug' 'denies etoh' 'denies']

3.4.4 *Scoring Results and Comparison*

After classifying text for housing stability, tobacco use, and alcohol use for patients in Group B, we applied a scoring metric scheme, described in the Methods section. We generated scores that were calculated and weighted differently based on time. Our final score weighs more recent note entries and their resulting classification score higher than notes from previous years as social factors and their influence can change over time. Using the same process, we extracted and scored housing stability, tobacco use, and alcohol use with structured data sources and compared the results with the unstructured process.

3.4.4.1 Housing stability

Using notes, we classified 839 patients as housing unstable, a score above 0.5, and 21,370 patients as housing stable, a score of 0.5 and below. In total, we classified 22,209 patients with this text classification workflow, which covered 50.7% of the acute care patients within the same timeframe. When compared with structured data sources, only 791 (1.8%) additional patients were found.

3.4.4.2 Tobacco Use

We classified 4,039 patients as currently using tobacco, regardless of amount or degree (2) using text notes. We classified 2,423 patients as having rare/occasional/past use of tobacco (0-2), and 6,492 patients as not using tobacco (0). In total, we classified 12,954 patients with this text classification workflow, which covered 28.7% of the acute care patients within the same timeframe. When compared with structured data sources, 18,730 (42.7%) additional patients were captured.

3.4.4.3 Alcohol Use

We classified 2,196 patients as currently using alcohol, regardless of amount or degree (2) using text notes. We classified 4,405 patients as having rare/occasional/past use of alcohol (0-2), and 9,612 patients as not drinking alcohol (0). In total, we classified 16,213 patients with this text classification workflow, which covered 37% of the acute care patients within the same timeframe. When compared with structured data sources, we found an additional 135 patients, which covered 0.3% of acute care patients.

3.4.4.4 Structured and Unstructured Data Comparison

Structured sources of patient social and health behavior information were captured in numerous locations within the electronic medical record, often times in overlapping areas. Structured housing information was pulled from the following locations: (1) Problem list (ICD 9), (2) Patient encounter homelessness question (Y/N), (3) diagnosis (ICD 9 or 10), Patient address (typically would state homeless, DESC, shelter, use the hospital's address, or a DESC or shelter address), Q&A form with a housing question (e.g. "Where did you sleep last night?"), SmartForm data, or Clinical Event Discharge Disposition Entry. Table 3.7 lists the addresses of the shelters that were queried. We found that unstructured data for housing stability identified a total of 17,511 total patients, which covered 40% of acute care patients. With structured data sources, we were able to identify an additional 1,093 patients that were not found with clinical notes, of which 2.5% were in acute care.

Structured data sources for alcohol use consisted of (1) Problem list (ICD 9), (2) Patient encounter (Y/N), (3) diagnosis (ICD 9 or 10), and (4) structured social history data entries. Clinical events were also explored but were too messy and could not be considered structured. We found 16,213 patients with unstructured text data, which covered 37% of acute care patients. With structured data sources, we found an additional 135 patients, which covered 0.31% of acute care patients.

Similarly, structured data sources for tobacco use consisted of (1) Problem list (ICD 9), (2) Patient encounter (Y/N), (3) diagnosis (ICD 9 or 10), and (4) Epic SmartForm. We found 12,954 patients with unstructured text data, which covered 28.7% of acute care patients.

However, with structured data sources, we found an additional 18,730 patients that were not found using text data. This covered 42.7% of the acute care patients.

Table 3.6. List of addresses used to identify patients experiencing housing instability

Harborview Medical Center	'325 9th Ave', '325 Ninth Ave'
Downtown Emergency Service Center (DESC) Facilities	Main Office: '515 3rd Ave', '515 Third Ave' Emergency Shelters: '517 3rd Ave', '517 Third Ave', '505 3rd Ave', '505 Third Ave', '510 Minor Ave N', '606 12th Ave S', '606 Twelfth Ave S' Supportive Housing Projects: '1811 Eastlake Ave', '10507 Aurora Ave N', '424 Minor Ave N', '937 N 96th St', '5444 Delridge Way SW', '3501 Rainier Ave S', '415 10th Ave', '415 Tenth Ave', '2208 15th Ave W', '510 Minor Ave N', '607 3rd Ave','607 Third Ave', '509 3rd Ave','509 Third Ave', '5270 Rainier Ave S'
Shelters in Seattle	'2030 3rd Avenue', '2720 E Madison St', '611 12th Ave S', '150 Denny Way', '600 4th Ave', '600 Fourth Ave', '1265 S Main St', '715 Spring St', '505 Third Avenue', '157 Roy St', '2810 E Cherry St', '302 N 78th St', '326 Ninth Ave', '326 9th Ave', '1215 Thomas St', '4th & Jefferson', '510 Minor Ave N', '500 4th Ave', '500 Fourth Ave', '3120 NE 125th St', '2329 Rainier Ave S',

	'2709 3rd Ave', '2709 Third Ave', '118 Bell St', '2030 Third Avenue', '2030 3rd Avenue', '901 Rainier Ave S', '1101 Pike St', '1609 19th Ave', '1415 NE 43rd St', '232 Warren Ave', '305 Harrison St', '301 Mercer St', '606 12th Ave S', '1561 Alaskan Way S', '318 2nd Ave', '318 Second Ave', '811 Maynard Ave S', '3800 S Myrtle St', '2100 24th Avenue S', 'Rainier Ave S & S Hudson St', 'Denny Way & Stewart St

3.5 Discussion

Our approach using a simple text classification method for our chosen social and behavioral determinants of health has shown positive results. The selected classification models were chosen as they were the most commonly used classification models when researching text classification techniques. Furthermore, these models were robust enough to curtail the need for more complex machine learning based text classification methods, which may be harder to interpret in the clinical space as the weights and decisions can cause confusion due to the black box nature of these more complex classification methods; furthermore, in the healthcare field, where there are many high-stakes situations, it may be better to use more interpretable methods.[34] Generally, linear models are fast to train, can work well with sparse data, and offer interpretability.[35] Additionally, recent research has also suggested that more complex machine learning approaches may not yield statistically significant improvements in predictive power to justify the time and effort necessary to implement and test these more complex methods.[36,37]

Although promising, more advanced methods of NLP, such as convoluted neural networks, may not provide a significant tradeoff in improvement or accuracy versus transparent understanding of rule-based approaches. In fact, Yao et al. found that the F1 scores for CNN via TensorFlow did not improve significantly for interested features when compared to logistic regression and support vector machine implementations.[36] Finally, generalizable methods to create institution-specific models can be better for the healthcare system as a whole since each institution collects, records, and stores clinical information differently.

Although SBDH information can be indicative of overall health, collection of SBDH heavily relies on clinical staff to screen and document SBDH. Furthermore, it also assumes that patients will respond accurately and truthfully. Various financial incentives from the federal level have propelled collection of social factors, such as tobacco use and tobacco cessation. However, other social factors, which can be equally as important, such as alcohol use are not incentivized to be captured; rather only more severe instances are incentivized, such as alcohol dependence or alcohol addiction or disorder.[38,39] Due to this discrepancy, we found that structured data sources were less reliable and less complete, and that text classification aided in detailing a patient more holistically, increasing the data completeness.

Our text classification of unstructured data relied solely on ED, admit, social work, and ambulatory notes. Social factors and other social history could also be recorded in other locations. Furthermore, social work and ambulatory notes used for housing status only and were only extracted if the notes contained a word or phrase related to housing instability. This approach was used as the notes were typically stored in a more unstructured format compared to the ED and admit notes; there were no section headers. The lack of section headers increased the difficulty to extract the notes and the notes would often verbiage that would interfere with the simple text classification approach that we used. Therefore, we decided to extract notes that contained words relating to housing instability. Additionally, tobacco and alcohol use notes had stylistic and grammatical challenges. These social factors were often grouped together in incomplete triples (e.g. “denies drinking, smoking, illicit drug use”). The classification algorithms often had trouble reciprocating the negative connotation to all components of the triple. Therefore, we used regex to specifically extract these triples and classify the note based on the presence of words related to tobacco or alcohol. These results would then override the text classification algorithm, if there was a discrepancy. Therefore, the scoring metrics for these cases would not necessarily reflect the accuracy or performance of our scoring method.

It was interesting to find that tobacco use was recorded significantly more often in structured data sources compared to alcohol use and housing stability. However, because tobacco use is a Centers for Medicare and Medicare Services (CMS) core quality measure, it can be expected that this feature is more available in structured form as it is often directly asked to the patient on intake forms, screeners, or during cessation treatment. Furthermore, the Joint Commission created the Tobacco Performance Measure Set, which are three standardized performance measures addressing tobacco screening and cessation counseling: (1) Tobacco use screening of patients 18 years and over, (2) Tobacco use treatment, including counseling and medication during hospitalization, and (3) Tobacco use treatment management plan at discharge. CMS began using these performance measures in 2016.[40]

3.5.1 *Limitations*

Our study has numerous limitations. There were two distinct areas in our workflow that required manual attention: (1) EHR review and (2) labelling of features. Manual EHR review was performed to ensure that the notes contained social history information in a consistent location prior to widespread text extraction. We initially validated this with a random set of 10 patients, but later expanded our validation to 25 patients. We felt that having consistent results with the 25 patients indicated a high level of confidence. Manual labelling of features was time consuming and taxing. Although only one author performed the feature labelling, having multiple team members would provide better and possibly more consistent classification.

This approach, although we aim to create a generalizable workflow, is still stunted by local customizations due to unique nuances in note taking language. Patients can withhold information about their social challenges, making text classification harder to perform due to incorrect incoming data streams. Our approach relies on the fact that the patient has been seen within the healthcare system at some point in the past five years. This approach would not be applicable to those who are new to the institution or those who are not immediately identifiable. Classification levels for unstructured notes are not concrete as descriptive wording is also not concrete and can vary (e.g. “patient was a former smoker”, “patient quit last week”, “patient is an occasional smoker”, etc.). Structured data sources can add a more concrete sense to the classification. There were 5.7% copy-forward entries present as data collection of social factors may not always be appropriate (e.g. patient is inebriated, in an altered mental state, etc.). We did not incorporate outside ontologies, such as UMLS or MetaMap, as we were interested in creating a simple text classification approach that did not need to rely on outside entities. Furthermore, we believe that these ontologies would not have added a significant improvement in our approach due to the social factors (housing, alcohol, tobacco) that were investigated. Although minimized, applying NLP to clinical notes will always present limitations and risks with biased models, biased data, and data privacy.[41]

3.6 CONCLUSION

From our analysis, we can first see that text classifiers are promising when applied to extracted clinical notes for housing stability, tobacco use, and alcohol use status. Additionally, we found that structured data sources, such as diagnosis codes and intake surveys, vary and may not be the most holistic approach to understanding housing stability, tobacco use, and alcohol use. Our simplified approach has shown that open source simple text classifiers can be used to predict text sentiment for social determinants and can supplement current structured sources to provide a more complete social history for patients. However, even with a few limitations with our approach, we believe that this workflow can help inform clinicians and provide an easily implementable snapshot on patient social history.

3.7 REFERENCES FOR CHAPTER 4

1. Adler-Milstein J, Jha AK. HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. *Health Affairs*. 2017;36(8):1416-1422. doi:10.1377/hlthaff.2016.1651

2. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. *Journal of the American Medical Informatics Association*. 2016;23(2):375-379. doi:10.1093/jamia/ocv103

3. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*. 2014;2(1). doi:10.1186/2047-2501-2-3

4. Structured and Unstructured Health Data: Challenges and Opportunities. Published October 29, 2018. https://healthsnap.io/structured-unstructured-health-data/

5. ICD-10-CM Official Guidelines for Coding and Reporting FY 2020. https://www.cdc.gov/nchs/data/icd/10cmguidelines-FY2020_final.pdf

6. Hood CM, Gennuso KP, Swain GR, Catlin BB. County Health Rankings. *American Journal of Preventive Medicine*. 2016;50(2):129-135. doi:10.1016/j.amepre.2015.08.024

7. Gottlieb LM, Tirozzi KJ, Manchanda R, Burns AR, Sandel MT. Moving Electronic Medical Records Upstream. *American Journal of Preventive Medicine*. 2015;48(2):215-218. doi:10.1016/j.amepre.2014.07.009

8. Social Determinants of Health. HealthyPeople.gov. Accessed February 1, 2020. https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health

9. Social Determinants. Institute for Health Metrics and Evaluation. Accessed February 1, 2020. http://www.healthdata.org/social-determinants

10. Nerenz DR. Health Care Organizations' Use Of Race/Ethnicity Data To Address Quality Disparities. *Health Affairs*. 2005;24(2):409-416. doi:10.1377/hlthaff.24.2.409

11. Andermann A. Taking action on the social determinants of health in clinical practice: a framework for health professionals. *Canadian Medical Association Journal*. 2016;188(17-18):E474-E483. doi:10.1503/cmaj.160177

12. Olson DP, Oldfield BJ, Navarro SM. Standardizing Social Determinants Of Health Assessments. Published March 18, 2019. https://www.healthaffairs.org/do/10.1377/hblog20190311.823116/full/

13. Wockenfuss R, Frese T, Herrmann K, Claussnitzer M, Sandholzer H. Three- and four-digit ICD-10 is not a reliable classification system in primary care. *Scand J Prim Health Care*. 2009;27(3):131-136. doi:10.1080/02813430903072215

14. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620-1639. doi:10.1111/j.1475-6773.2005.00444.x

15. Lax Y, Martinez M, Brown NM. Social Determinants of Health and Hospital Readmission. *Pediatrics*. 2017;140(5):e20171427. doi:10.1542/peds.2017-1427

16. King County Community Health Needs Assessment 2018/2019. Presented at the: https://www.kingcounty.gov/depts/health/data/community-health-indicators/~/media/depts/health/data/documents/2018-2019-Joint-CHNA-Report.ashx

17. Henry M, Mahathey A, Morrill T, et al. *The 2018 Annual Homeless Assessment Report (AHAR) to Congress*. The U.S. Department of Housing and Urban Development OFFICE OF COMMUNITY PLANNING AND DEVELOPMENT; 2018. https://files.hudexchange.info/resources/documents/2018-AHAR-Part-1.pdf

18. Stafford A, Wood L. Tackling Health Disparities for People Who Are Homeless? Start with Social Determinants. *International Journal of Environmental Research and Public Health*. 2017;14(12):1535. doi:10.3390/ijerph14121535

19. Papadopoulou S, Hassapidou M, Katsiki N, et al. Relationships Between Alcohol Consumption, Smoking Status and Food Habits in Greek Adolescents. Vascular Implications for the Future. *Current Vascular Pharmacology*. 2017;15(2):167-173. doi:10.2174/1570161114666161024123357

20. Wong E. *Tobacco Use in King County*. Public Health Seattle & King County; 2012. https://www.kingcounty.gov/depts/health/data/~/media/depts/health/data/documents/tobacco-use-in-king-county-may-2012.ashx

21. Bogan S, Donohue B. King County drug and alcohol deaths rose 9.5% in 2018. https://newsroom.uw.edu/news/king-county-drug-and-alcohol-deaths-rose-95-2018

22. Drug-caused deaths in King County. Published online February 21, 2017. https://adai.washington.edu/WAdata/KingCountyDrugDeaths.htm

23. Gundlapalli AV, Carter ME, Palmer M, et al. Using natural language processing on the free text of clinical documents to screen for evidence of homelessness among US veterans. *AMIA Annu Symp Proc*. 2013;2013:537-546.

24. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010;17(5):507-513. doi:10.1136/jamia.2009.001560

25. Gundlapalli AV, Carter ME, Divita G, et al. Extracting Concepts Related to Homelessness from the Free Text of VA Electronic Medical Records. *AMIA Annu Symp Proc*. 2014;2014:589-598.

26. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS ONE*. Published online 2017. doi:10.1371/journal.pone.0174708

27. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using Clinical Notes and Natural Language Processing for Automated HIV Risk Assessment: *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2018;77(2):160-166. doi:10.1097/QAI.0000000000001580

28. Dorr D, Bejan CA, Pizzimenti C, Singh S, Storer M, Quinones A. Identifying Patients with Significant Problems Related to Social Determinants of Health with Natural Language Processing. *Stud Health Technol Inform*. 2019;264:1456-1457. doi:10.3233/SHTI190482

29. Berg K, Doktorchik C, Quan H, Saini V. *Meaningful Information in the Age of Big Data: A Scoping Review on Social Determinants of Health Data Collection for Electronic Health Records*. In Review; 2019. doi:10.21203/rs.2.16433/v1

30. Ahmad S, Baig S, Taneja A, Nanchal R, Kumar G. The Outcomes of Severe Sepsis in Homeless. *Chest*. 2014;146(4):230A. doi:10.1378/chest.1995140

31. Bambra C, Gibson M, Sowden A, Wright K, Whitehead M, Petticrew M. Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. *Journal of Epidemiology & Community Health*. 2010;64(4):284-291. doi:10.1136/jech.2008.082743

32. 2015 CDC HA-VTE PREVENTION CHALLENGE CHAMPION. https://www.cdc.gov/ncbddd/dvt/documents/champ-fact-sheet-harborview.pdf

33. Bulger EM, Kastl JG, Maier RV. The history of Harborview Medical Center and the Washington State Trauma System. *Trauma Surgery & Acute Care Open*. 2017;2(1):e000091. doi:10.1136/tsaco-2017-000091

34. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206-215. doi:10.1038/s42256-019-0048-x

35. Cronin RM, Fabbri D, Denny JC, Rosenbloom ST, Jackson GP. A comparison of rule-based and machine learning approaches for classifying patient portal messages. *International Journal of Medical Informatics*. 2017;105:110-120. doi:10.1016/j.ijmedinf.2017.06.004

36. Yao L, Mao C, Luo Y. Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. *BMC Medical Informatics and Decision Making*. 2019;19(S3). doi:10.1186/s12911-019-0781-4

37. Chicco D, Rovelli C. Computational prediction of diagnosis and feature selection on mesothelioma patient health records. Vellido A, ed. *PLoS ONE*. 2019;14(1):e0208737. doi:10.1371/journal.pone.0208737

38. Medicare & Medicaid EHR Incentive Program: Meaningful Use Stage 1 Requirements Overview. Presented at the: 2010. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/MU_Stage1_ReqOverview.pdf

39. Eligible Professional Meaningful Use Core Measures Measure 9 of 13. Published online May 2014. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/9_Record_Smoking_Status.pdf

40. Quality Measures and Tobacco Cessation. https://www.bhthechange.org/wp-content/uploads/2017/12/Quality-Measures-and-Tobacco-Cessation.pdf

41. Baclic O, Tunis M, Young K, Doan C, Swerdfeger H. Challenges and opportunities for public health made possible by advances in natural language processing. *Canada Communicable Disease Report*. Published online June 4, 2020:161-168. doi:10.14745/ccdr.v46i06a02

Chapter 4. INTEGRATING SOCIAL AND HEALTH BEHAVIOR DATA WITH ESTABLISHED CLINICAL SCORES FOR POTENTIAL SEPSIS ACUTE CARE PATIENTS

4.1 ABSTRACT

Background: Detecting sepsis has been a challenge as preliminary symptoms are often not specific and similar to other diseases. Identifying important features to extract to aid in sepsis identification has resulted in numerous definitions. However, all established rules and scoring schemes focus on vital signs and lab results. Social determinants also have been shown to have a correlation with health and clinical outcomes.

Objective: To explore how social determinants of health data extracted from unstructured clinical notes can impact clinical scoring schemes commonly used for potential sepsis patients in acute care.

Methods: We integrated results from an exploration that extracted social determinants of health data from unstructured clinical notes and created scores for housing stability, tobacco use, and alcohol use. These scores were generated using a longitudinal approach and a prior-year approach. These scores were then integrated with extracted vital signs and lab results. We examined the influence of these additional social features by comparing baseline SIRS, qSOFA, and MEWS scores with their enhanced counterparts. We identified sepsis patients as having a diagnosis code related to sepsis, a clinical event where an antibiotic was administered, and data from a nursing checklist that indicated whether the nurse suspects signs of infection or were already treating the patient for sepsis.

Results: From our analysis, we found that the addition and integration of our selected social and health behavior features did not necessarily influence the existing scoring schemes commonly used for potentially sepsis patients. With the longitudinal approach, we found that the baseline accuracies were 0.66 for qSOFA, 0.71 for SIRS, and 0.63 for MEWS. The prognostic accuracies with social and health behavior information were 0.68 for qSOFA, 0.67 for SIRS, and 0.63 for MEWS. With the prior year approach, we found the baseline accuracies were 0.64 for qSOFA, 0.68 for SIRS, and 0.63 for MEWS. The prognostic accuracies with social and health behavior information were 0.56 for qSOFA, SIRS, and MEWS.

Keywords: Sepsis detection, Social determinants, Health behavior, Acute care

4.2 BACKGROUND

According to the Centers for Disease Control and Prevention (CDC), approximately one third of patients who die in a hospital had sepsis, which is an extreme reaction to an infection that can lead to tissue damage, organ failure, or even death.[1] Because those with sepsis have nondescript early onset symptoms, detection and diagnosis of sepsis have been challenging.[2] The importance of detecting sepsis in a timely manner is apparent as sepsis can quickly progress to a life-threatening stage and is a strain on health care systems due to is high prevalence and high cost.[3]

Throughout the past decades, various rule-based definitions and scoring schemes were created to help define and identify sepsis in the clinical setting. First in 1991, the American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference introduced the clinical definitions of sepsis, severe sepsis, and septic shock, commonly referred together as Sepsis-1.[4] Sepsis-1 defined sepsis as a systemic inflammatory response syndrome (SIRS) due to a present infection, with at least two of the following criteria: (1) Temperature greater than 38 degrees Celsius or less than 36 degrees Celsius, (2) Heart rate greater than 90 beats per minute, (3) Respiratory rate greater than 20 per minute or $PaCO_2$ less than 32 mmHg, or (4) White blood cell count greater than 12,000 per mm^3, less than 4,000 per mm^3, or greater than 10 percent bands; severe sepsis was having sepsis resulting in organ dysfunction, hypoperfusion, or hypotension while septic shock was the occurrence of sepsis-induced hypotension despite adequate fluids.[5,6] Because SIRS is nonspecific, the presence of at least two SIRS criteria may not always be the result of an infection. One decade later, the International Sepsis Definitions Conference updated Sepsis-1 with Sepsis-2, which added confirmed or suspected infection to the sepsis definition.[5] However, in 2016, Singer et al. created Sepsis-3, ultimately changing the definitions of sepsis and septic shock that were in use for more than two decades.[6] Under Sepsis-3, sepsis is now described as a life-threatening organ dysfunction caused by a dysregulated host to infection and there are currently two scoring schemes available for determining organ dysfunction: (1) Sequential Organ Failure Assessment (SOFA) Score to calculate organ dysfunction related to sepsis, often used in the ICU setting, and quickSOFA (qSOFA) to identify patients with high risk of adverse outcomes, often used in the non-ICU setting.[6] SOFA takes various organ systems into account, including the respiratory, hematologic, hepatic, cardiovascular, neurologic, and renal systems.[6] On the other hand, qSOFA criteria consists of (1) respiratory rate $\geq$ 22 per minute, (2) a change in mental status, and (3) systolic blood pressure $\leq$ 100mmHg.[6] Outside the United States, detection of sepsis is also challenging. In 1999, the Audit Commission recommended developing early warning systems (EWSs) to help clinical staff identify patients who needed attention. They proposed a modification of Morgan's Early Warning Score, creating the Modified Early Warning Score (MEWS) which aimed to create communication between nursing and medical staff when patient deterioration was detected.[7] MEWS consists of the following clinical features: (1) systolic blood pressure, (2) heart rate, (3) respiratory rate, (4) temperature, and (5) Alert-Verbal-Pain-Unresponsive (AVPU) score. These rule-based definitions have been applied in the clinical setting to assist with sepsis detection with varying success.[8–11]

In addition to using established rule based clinical scores, Harborview Medical Center (HMC) developed a systematic rule-based screening tool in 2012 to identify sepsis inpatients using EHR data to evaluate patients for signs of infection for patients in acute care. The screening system utilized SIRS criteria as well as lactate and would notify providers if nurses suspected new or worsening infection. Internal pilot testing showed that rapid response team activations were reduced by half on the acute care floor. Additionally, it was shown that time from sepsis screen to care was reduced and 3-hour bundle compliance increased over a year. Furthermore, the average nurse response time to screen decreased.[12,13]

In addition to structured data and commonly used vital signs and lab results, social and behavioral determinants of health (SBDH) are now of interest to many researchers as recent literature has indicated that there is a correlation between SBDH and health outcomes.[14] Social determinants of health (SDoH) can include housing stability, access jobs and health care services, education level, language, and socioeconomic conditions.[15] These indicators are descriptors of populations which are useful health outcomes predictors.[16] Because SDoH can potentially be powerful indicators of health, many institutions are analyzing SDoH information, whether through clinical text notes or standardized coding, such as International Classification of Diseases (ICD) or other structured data sources. Additionally, health behaviors (tobacco and alcohol use), although not directly considered a SDoH but together form SBDH, can impact health decisions and outcomes. For example, one study found that participants who drank alcohol and reported tobacco use consumed more foods higher in fat and sugar, low in vitamins and minerals as well as foods, considered by them to be less healthy and prepared in a less healthy way.[17] Therefore, it may be important to look at health habits and SDoH together.[17]

Although there have been inroads in exploring how predictive analytics can be used for sepsis prediction, most studies have focused on patients in the intensive care unit (ICU) or emergency department (ED) and most have focused on applying SIRS, Sepsis-2 or Sepsis-3 definitions as gold standards. However, to our knowledge, little has been investigated in the acute care setting; which we defined as consisting of patients who have been admitted to the hospital for a stay that is longer than 24 hours and do not require critical or intensive care. We explored integrating SBDH with existing sepsis definitions in the acute care setting at HMC. Through use of supervised algorithmic-based text and sentiment classification machine learning methods on clinical text notes, we explored and integrated social features, which were collected in unstructured formats at varying points in a patient's stay, with common sepsis clinical scoring and prediction methods and analyzed their impact on predictive analytics for sepsis patients in acute care.

4.3 METHODS

4.3.1 *Study Design, Setting, and Eligibility*

This was a retrospective analysis of data collected between January 1, 2015 and December 31, 2019 from a level one trauma urban public hospital. All patients who were in the acute care setting were included in our initial query, which extracted various vital signs and lab results, including respiratory rate, temperature, heart rate, oxygen saturation (SpO2), systolic blood pressure, diastolic blood pressure, white blood cell count, and Richmond Agitation Sedation Scale (RASS). These vital signs and lab results were selected as they are necessary components for computing clinical scores commonly used for sepsis patients. Patient encounters that did not have at least one of these data points present were dropped.

4.3.2 *Data Collection and Quality Control*

All data were queried directly from Microsoft SQL Server Management Studio (SSMS), which contained patient data from both Cerner and Epic electronic medical record systems. The institution utilized Cerner Millennium for inpatient data collection. Vital signs and clinical notes were queried from Cerner clinical events while lab results and structured data (diagnosis or billing codes and ICD descriptors) were queried from Epic Clarity tables. An extract, transform, and load (ETL) process was performed to convert the varying data store formats into a common data model. The Observational Medical Outcomes Partnership (OMOP) common data model was chosen for its wider use and acceptance within the field.[18–22] All the data were averaged into one-hour time buckets.

Additionally, clinical notes were extracted for patients, not solely limited to acute care, to identify and score social and health behavior factors, including housing stability, tobacco use, and alcohol use. These generated scores were merged via an inner join with our acute care patient data by patient identifier, date, and hour. Further details on how these text notes were extracted and scored are in detailed in Chapter 3.

Integration of these scores were performed in two manners, longitudinal and prior year average. The longitudinal method joined the vital signs and lab results with social factor scores by encounter identifier. This kept the longitudinal recording of the social factors in notes data intact. For example, if a patient had a note detailing their housing status on January 1, 2015 at 08:00, then the housing score would be merged with the vital signs and lab results available at that exact day and hour if the encounter identifier was the same. On the other hand, the prior year average method took the scores for any given patient from the prior year, if available, and calculated a mean value which was then merged by patient identifier. Therefore, if a patient was seen in 2015 and had a housing score generated from the available notes from 2015, if the patient had any acute care encounter again in 2016, the housing score generated by the prior-year average would be joined with the available vital signs and lab results for the entire latter year, regardless of encounter identifier.

4.3.3 *Defining Sepsis*

We evaluated our entire study population for the presence of sepsis, suspected sepsis, or administration of an antibiotic within the patient encounter. A combination of ICD 9 or 10 codes along with a diagnosis description string matching query using sepsis related words, including “sepsis”, “bacteremia”, and “urosepsis”, were used to identify patients who had sepsis. Additionally, we specifically extracted the administration of an antibiotic regardless of administration time. A full list of the antibiotics that were included in our query can be seen in Table A4.1. Because we did not have expert review of the medical records or patient charts, in an effort to include and integrate more skilled knowledge in our definition of sepsis, we included binary data from the screening system described earlier which allowed nurses to indicate if they suspected new or worsening infection. This checklist included binary questions or entries like “Sepsis - Being treated for Infection (Y/N)”, Sepsis Screen - Suspected Infection (Y/N)”. The presence of an antibiotic and a diagnosis code, and a confirmation from the nursing checklist for either question formed our definition sepsis to identify septic patients.

4.3.4 *Sepsis and Scoring Systems*

Our target measure was the ability for the scoring systems to identify an instance of sepsis and integrate previously generated social factor and health behavior scores and explore how the addition of these scores can impact the performance of these existing sepsis or clinical deterioration scoring systems. Three scoring systems were selected for our study, (1) SIRS, (2) qSOFA, and (3) MEWS. Other existing scoring systems, like SOFA, were not used as they were not appropriate for our requirement of acute care patients only. Additionally, these scoring systems would require various other parameters and/or lab tests that would not be traditionally administered in this setting making these scoring systems difficult to use and integrate. Missing data were imputed via forward fill by patient identifiers and encounter identifiers for every available hour within the encounter. This preserved the existing process that nurses and physicians used by looking at prior encounters or charts to find the most recent data point. Imputing by date and time would have prevented data from 23:00 being forward filled for 01:00 the next day. Using encounter identifiers also helped with copy forwarding data points that were between days. The AVPU (Alert, Voice, Pain, Unresponsive) scale was a required component to qSOFA; however, this scale was not used at our institution. Rather we used the RASS and converted it to equivalent points on the AVPU scale, as seen in Table 4.1. Additionally, for each scoring system, a subset of the study population was used as not all hours of every encounters had the necessary features for calculating scores for each scoring system.

Table 4.1. RASS to AVPU Conversion

RASS	**AVPU**	**RASS Description**
Above 0	0 (Alert)	Alert, but agitated or restless
0 to 4	0 (Alert)	Alert and calm
-1 to -3	1 (Verbal)	Drowsy, light/moderate sedation
-4	2 (Pain)	Deep sedation, but response to physical stimulation
-5	3 (Unresponsive)	Unarousable

4.3.5 *Statistical analysis*

Data analysis was performed in Python through a Jupyter notebook, both open-source resources commonly used in the data science field. The SciKit learn package was utilized as it has already built machine learning and statistical functions. Comparisons between the original scoring schemes and the enhanced scoring schemes with social factors were performed using chi-squared tests to determine the relationship between the established scoring schemes and our integrated approach with social determinants. Accuracy, sensitivity, and specificity metrics were also calculated from a subset of the scored data, utilizing and filtering only the first hour bin of data where the clinical score was highest per encounter identifier. This was done to reduce redundancies and to minimize bias from any particular patient or encounter.

4.4 RESULTS

4.4.1 *Study population*

The demographics of our selected acute care population are detailed in Table 4.2. These patients had at least one of the required vital signs and lab results necessary to calculate qSOFA, SIRS, or NEWS. Our string-based query to identify patients with sepsis, bacteremia, or urosepsis identified 4,692 patients (5,566 unique encounters) with a sepsis related diagnosis. We also identified 32,830 patients (42,803 unique encounters) who had antibiotics administered at any given point during within the encounter. Using the two types of nurse checklist data points, we identified 18,690 patients (23,168 unique encounters) where the nurse suspected sepsis and 5,728 patients (6,590 unique encounters) where the nurse was already treating the patient for sepsis. Once the social factors were merged and integrated with this population subset, we had 28,387 patients (35,226 unique encounters) with housing data, 18,175 patients (20,883 unique encounters) with alcohol data, and 15,155 patients (17,005 unique encounters) with tobacco data all from extracted via text classification methods of unstructured data sources, clinical (ED and admission) notes.

During the 2015-2019 timeframe, we extracted 55,067 patients (75,749 unique encounters) were in acute care after extraction and removing patients who did not have an encounter identifier and those who were missing at least one vital sign or lab result. The patient demographics of this final subset were 61.9% (*n*=34,099) male, 38% (*n*=20,937) female, 85.7% (*n*=47,209) not Hispanic or Latino, and 9.4% (*n*=5,161) Hispanic or Latino, and 4.9% (*n*=2,697) unknown or not answered.

Table 4.2. Population demographics

Sex (*n*=55,067)	*n* (%)
Male	34,099 (61.9%)
Female	20,937 (38%)
Unknown	31 (0.1%)
Race (*n*=55,067)	*n* (%)
White or Caucasian	40,227 (73.1%)
Black or African American	5,573 (10.1%)
Asian	3,708 (6.7%)
American Indian or Alaska Native	1,238 (2.2%)
Native Hawaiian or other Pacific Islander	578 (1%)
Mexican, Mexican-American or Chicano	298 (0.5%)
Hispanic	304 (0.6%)
Middle Eastern	28 (0.05%)
Puerto Rican	11 (0.02%)
Multiple races	9 (0.02%)
Unavailable, unknown, or missing	3,093 (5.6%)
Age range (*n*=55,067)	*n* (%)
0-18	2,020 (3.7%)
19-44	14,767 (26.8%)
45-64	17,927 (32.6%)
65-84	16,147 (29.3%)
85 and over	4,206 (7.6%)

4.4.2 *Data Attributes*

Because each established scoring system uses different features for calculation, a subset of the prior described patient population was used for each scoring system to maintain data completeness for each subset. Therefore, encounters with a missing Glasgow coma score were not used for qSOFA calculation as the data was not present, even after imputing via forward fill by encounter identifier. Outcome data (presence of antibiotics, diagnosis, nursing suspicion or confirmation) and social history data (housing, alcohol, tobacco status) were converted to binary sets and missing data were replaced with zeros.

After removing rows that were missing data for each scoring subset, there were 53,199 total patients (73,374 unique encounters) that were used to calculate qSOFA, 28,362 total patients (35,852 unique encounters) for SIRS, and 14,062 total patients (15,452 unique encounters) for MEWS. Among those with an adjusted qSOFA score greater than or equal to 2, 2,935 out of 30,264 (9.70 %) patients had indicators of housing instability, 3,672 (12.13%) patients had indicators of tobacco use, and 3,408 (11.26%) patients had indicators of alcohol use all scored from extracted notes using natural language processing classification methods described in Chapter 3. Similarly, among those with an adjusted SIRS score greater than or equal to 2, 2,940 out of 28,362 (10.4%) patients had indicators of housing instability, 3,798 (13.4%) patients had indicators of tobacco use, and 3,590 (12.7%) patients had indicators of alcohol; and among those with an adjusted MEWS score greater than or equal to 2, 1,264 out of 14,062 (9.0%) patients had indicators of housing instability, 1,576 (11.2%) patients had indicators of tobacco use, and 1,485 (10.6%) patients had indicators of alcohol. These cutoffs were used as they are standard in the clinical setting for determining clinical deterioration.

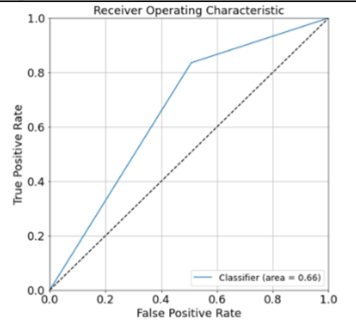
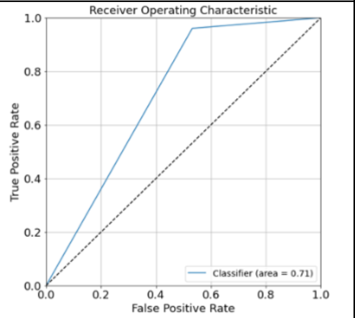
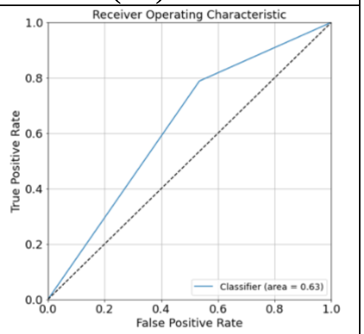
4.4.3 *Model Performance*

4.4.3.1 Longitudinal Approach

Using our filtered data, our definition of sepsis to identify septic patients (presence of antibiotics, diagnosis/billing, and nurse suspicion/nurse confirmation) was compared with the scores created by qSOFA, SIRS, and MEWS. The AUROC for the detection of sepsis at baseline were qSOFA = 0.66, SIRS = 0.71, and MEWS = 0.63. The AUROC for the detection of sepsis with SBDH scores averaged together were qSOFA = 0.68, SIRS = 0.67 and MEWS = 0.63. The AUROC for the detection of sepsis with just averaged SDoH was 0.57 with qSOFA features, 0.52 with SIRS features, and 0.55 with MEWS features. The accuracy of these comparisons can be seen in Table 4.3. Area under the characteristic receiver operator curve plots can be seen in Figure 4.1. Furthermore, the integrated social determinants data with these scoring schemes and the results of this integration can be seen in Table 4.3.

Table 4.3. Accuracies of clinical scoring schemes and SBDH – longitudinal approach

	qSOFA (2+) filtered	SIRS (2+) filtered	MEWS (2+) filtered
Baseline	0.66	0.71	0.63
Scores with social and health behavior	0.68	0.67	0.63

	qSOFA (2+) filtered	SIRS (2+) filtered	MEWS (2+) filtered
Baseline			

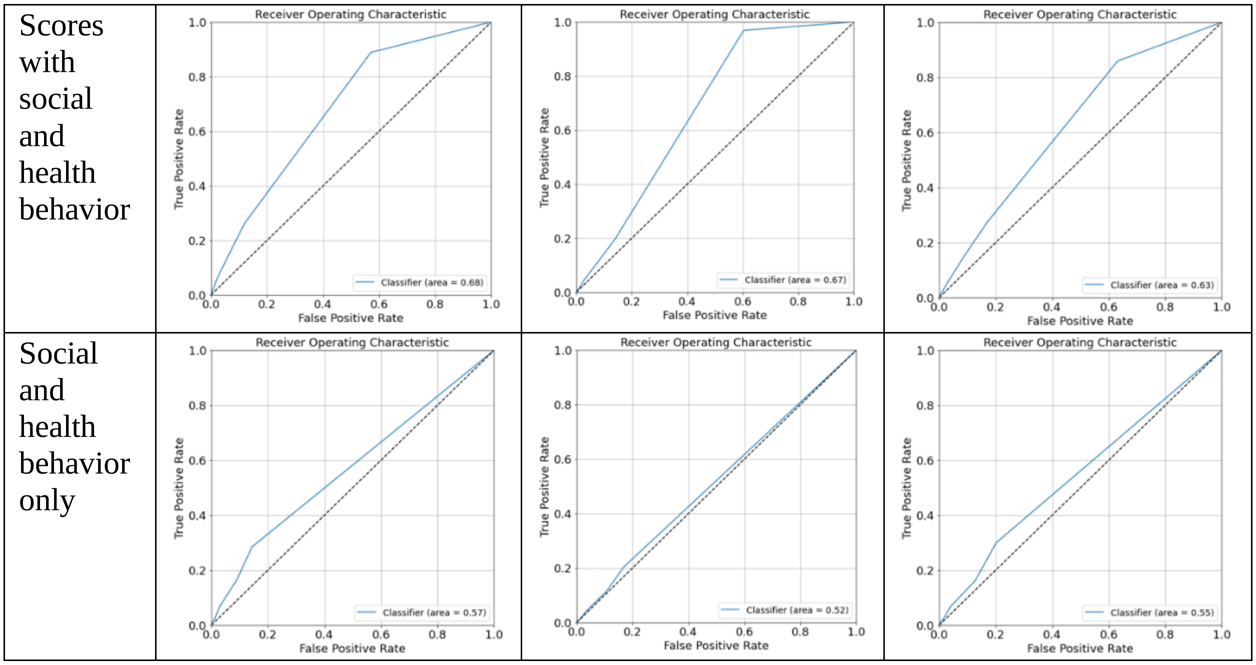


Figure 4.1. Area under the characteristic receiver operator curve plots – longitudinal approach.

4.4.3.2 Prior Year Approach

Using social history and health behavior information from the previous year of the encounter was also used. If a patient was seen in the year prior the scores generated within that year for social determinants were averaged over the year and integrated with the current encounter. The AUROC for the detection of sepsis were qSOFA = 0.64, SIRS = 0.58, and MEWS = 0.63. The AUROC for the detection of sepsis with SDoH scores were qSOFA = 0.56, SIRS = 0.56 and MEWS = 0.56. The AUROC for the detection of sepsis with just SDoH was 0.50 with qSOFA features, 0.48 with SIRS features, and 0.51 with MEWS features. Similarly, more details can be seen in Table 4.4 and Figure 4.2.

Table 4.4. Accuracies of clinical scoring schemes and social and health behavior – prior year approach

	qSOFA (2+) filtered	SIRS (2+) filtered	MEWS (2+) filtered
Baseline	0.64	0.68	0.63
Scores with social and health behavior	0.56	0.56	0.56

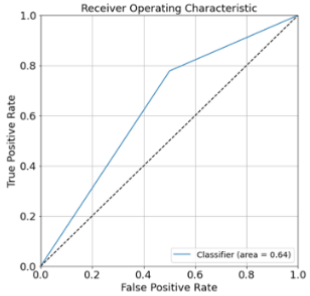
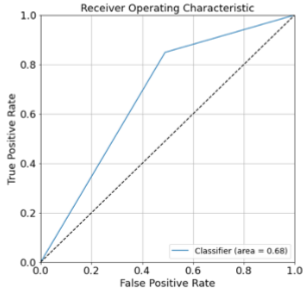
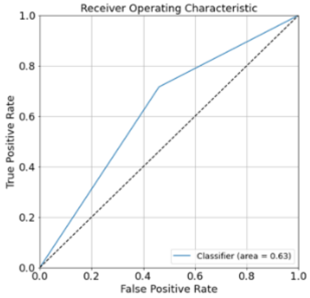
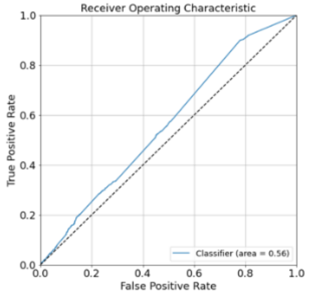
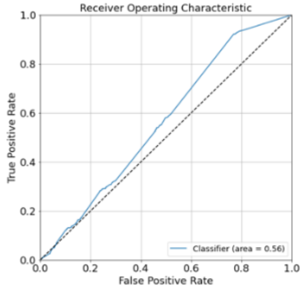
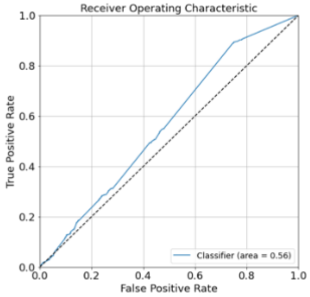
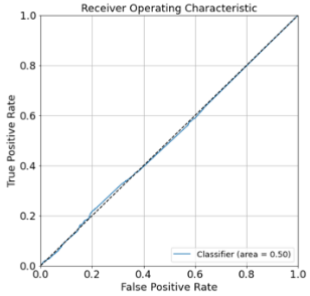
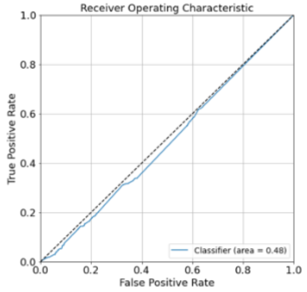
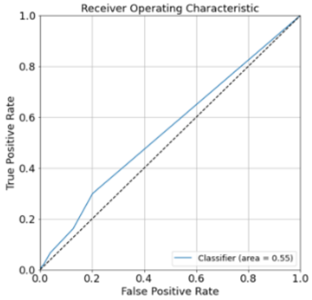
	qSOFA (2+) filtered	SIRS (2+) filtered	MEWS (2+) filtered
Baseline			
Scores with social and health behavior			
Social and health behavior only			

Figure 4.2. Area under the characteristic receiver operator curve plots – Prior year approach.

Significance

We used a chi-squared test to determine whether there was a relationship between the scoring schemes (qSOFA, SIRS, MEWS) and our definition of sepsis which identified septic patients. Afterwards, we sought to determine whether there was a relationship between the scoring schemes integrated with SBDH for sepsis patients. The results of this analysis can be seen in Table 4.5. Sensitivity and specificity calculations were only performed on the baseline scores as scores with social determinants created continuous, non-integer scores as they were averaged together. For the unfiltered scores, meaning the data sets where the maximum score per encounter identifier was not singled out, we can see across the board, all scores have very low sensitivity and relatively higher specificity. There was also poor association with likelihood and sepsis with the three scoring methods.

When filtering out for max score value per encounter identifier, we saw a reversal in sensitivity and specificity. In these cases, there was high sensitivity and low specificity, meaning it did well identifying those who were likely to develop sepsis, but produced a high false-positive rate. With this subset, there was a stronger association between our sepsis definition and the selected scoring methods. When narrowing down to each targeted SBDH feature with the filtered data subsets, we found that housing and tobacco had associations with the scoring methods and their detection of sepsis likely; however, alcohol use was a poor indicator. Furthermore, a feature importance was calculated on the scoring methods to see which vital sign or lab result influenced the scores the most; however, no clear pattern of importance arose or could be identified from the analysis.

Table 4.5. Sensitivity and Specificity for baseline scores

	qSOFA (2+) unfiltered	SIRS (2+) unfiltered	MEWS (2+) unfiltered	qSOFA (2+) filtered	SIRS (2+) filtered	MEWS (2+) filtered
Sensitivity	0.35	0.27	0.14	0.84	0.96	0.79
Specificity	0.66	0.81	0.88	0.49	0.47	0.46
χ^2, p-value (random 5K)	2.12, 0.146	24.71, 6.67×10^{-7}	0.36, 0.55	102.16, 5.11×10^{-24}	115.47, 6.21×10^{-27}	153.31, 3.28×10^{-35}

(random 5K)	qSOFA filtered	SIRS filtered	MEWS filtered
Housing: χ^2, p-value	49.90, 1.61×10^{-12}	19.18, 1.19×10^{-5}	36.07, 1.90×10^{-9}
Alcohol: χ^2, p-value	0.26, 0.88	0.23, 0.89	2.23, 0.33
Tobacco: χ^2, p-value	9.06, 0.011	8.57, 0.014	7.59, 0.022

4.5 DISCUSSION

At baseline, we found that SIRS performed more accurately compared to qSOFA and MEWS for identification of potential sepsis retrospectively in the acute care setting. When integrating social determinant and health behavior data scores, we found that qSOFA performed slightly more accurately compared to SIRS and MEWS for identification of sepsis retrospectively in the acute care environment. However, because we maximized the amount of data per scoring method by only excluding parts of an encounter that were missing required features per scoring method, the number of encounters and unique patients varied slightly across the board, making it harder to directly compare performance. Additionally, to reduce bias from any particular patient, we selected the first highest hourly bin per encounter identifier; doing so drastically shifted the distribution of the outcomes to skew more towards a 50/50 distribution for the clinical scores. However, the integration of social factors or health behavior data from unstructured clinical notes did not improve accuracy or performance in both a longitudinal approach or prior year approach. In fact, the use of housing stability, tobacco use, or alcohol use lowered performance overall when integrated with existing clinical scoring schemes and using purely social factors or health behavior did not correlate well with our definition of likely having sepsis in the acute care setting.

4.5.1 *Clinical Scoring Schemes*

qSOFA was designed to aid in identifying patients with suspected infection who are at a great risk of poorer outcomes in settings outside of the ICU. Previous literature has found that qSOFA has a high specificity and low sensitivity for determining patient mortality from sepsis.[23–25] This matches our initial results; however, this does not match our findings after we took a subset of the first maximum per encounter, which resulted in the exact opposite. This also may be due impart of our simplified definition to identify sepsis patients, rather than death from sepsis or even due to the fact we did not separate or aggregate out various sepsis severity levels. Additionally, because qSOFA requires the GCS, it may not be feasible to calculate for all patients in acute care as not all patients receive a neurological assessment. Additionally, qSOFA may not be as clinically sensitive as it does not have many critical physiologic criteria, such as respiratory rate, heart rate or temperature.[26,27] Other have suggested that qSOFA be used for monitoring purposes or for early identification of an infection or clinical deterioration. Interestingly, this matches our definition.[6,28]

SIRS was used as part of the original Sepsis-1 definition that was created about three decades ago. Previous literature has shown SIRS to be sensitive but not specific for sepsis detection due to the components that constitute SIRS and how they can be indicators for many other illnesses.[6,29] Our findings have reflected existing trends. Because MEWS is a rule-based early warning score, it is best used to identify patients who need immediate medical attention. It is superior to other rule-based systems as it does not require lab results to calculate the score and can be performed rather immediately, making it an ideal scoring scheme. In our case, we were classifying patients who would be "sepsis likely". At a threshold of 4, MEWS was previously found to be 75% sensitive for patients being admitted to a critical care facility with a specificity of 83%. However, in our implementation, we found that MEWS resulted in only a 45% specificity and 79% sensitivity. Within the ICU setting, Khwannitmit et al. found that early warning scores (MEWS and NEWS) were sufficient alterative tools to use for risk stratification and sepsis screening if SOFA scores were unable to be calculated.[11]

Interestingly, our outputs for the AUROC presented rather unsmooth curves. This can most likely be attributed to the fact that our scores are discrete and not continuous. Our data types are all dichotomous and categorical. Additionally, if we were to add more samples to the dataset, the curves may be smoother. Potentially using probabilities of the predicted class from the clinical scores and the observed labels from the "gold definition" could smooth the curve.

Furthermore, using cross validation techniques, such as leave-one-out, could also smoothen the curve for all the clinical scores.

4.5.2 *Limitations*

4.5.2.1 Sepsis Definition

Our method of defining a sepsis patient utilized underlying principles commonly used to identify a sepsis patient such as presence of an ICD code, infection suspicion, and administration of antibiotics. The type of data that we were interested in exploring did not necessitate the need to aggregate the varying levels of severity. Additionally, because we were interested in investigating how SBDH data would affect the scoring capabilities of existing scoring systems. Additionally, we were not calculating scores as an early-warning but rather taking the highest calculated score during an encounter and comparing it directly with our definition of identifying sepsis patients. Another difference was that ICD codes were not properly encoded in the database under the diagnosis table; however, were correctly encoded in the billing table. We had to approach the diagnosis table via a string search which potentially miss instances that did not explicitly mention sepsis, septicemia, or urosepsis in the diagnosis description.

4.5.2.2 Setting

Performance in the acute care setting overall was lower when compared to other hospital setting like the ED or ICU.[30–34] Additionally, our method of querying and extracting patients in the acute care setting was a bit unique. Studies in the existing literature tend to identify and include patients based off of their location at a given time and then start the filtering and exclusion process. However, we queried for vital signs and laboratory results first then filtered by patient location. Additionally, scoring information for the social concepts were performed independently from this query and were merged later by patient identifier, data, and hour. This can potentially lead to data that is available to be not merged as vital signs and laboratory results may not occur at the same hour as the intake note. In cases like this, the social features were therefore not merged as there was no equivalent hour to merge on. To facilitate comparison between our results and the existing literature, we also ran our queries for those in the ICU and the results were not much better. The full results of this comparison can be seen in Table A4.1.

4.5.2.3 Social Determinants and Health Behavior Features

Our analysis only utilizes social determinants data that was extracted from ED and admission notes. Integrating these unstructured sources of data with structured sources may improve correlation between SBDH and existing sepsis scoring systems. Additionally, our definition of sepsis could be improved upon by integrating professional and clinical opinion via chart reviews.

We found that the integration of social factors with existing sepsis scoring schemes did not improve predictive performance across the board. Although the scoring of social factors can provide clinicians an immediate numerical value to indicate recent housing instability, tobacco, or alcohol use, the inclusion of these factors ultimately did not boost accuracy when combined with vital signs and lab tests. Additionally, our target population was neither the ED or ICU, which are the two main areas of research for sepsis prediction, but rather the acute care setting. Patients in acute care would not necessarily have the same level of monitoring and testing done as those in the ICU. Therefore, certain scoring schemes, such as SOFA, were not tested due to the limited amount of potential data.

4.5.2.4 Expert Data: Nursing Checklist

The nursing checklist was designed to capture both binary and qualitative data from nurses who suspected an infection. Due to the complex nature of notes, these data were not utilized but could potentially provide another layer of expert data. For the system, an alert will sound once every 12 hours to prevent alert fatigue amongst the staff. However, it is possible for nurses to suspect new or a worsening infection and begin to treat the patient before an alert is generated. These cases would not be captured in the database. Additionally, although the alerts are derived from a simple rule-based system, there is a delay in creating these alerts as it is not directly integrated with the EHR and manual entry of the lab results and vital signs are necessary. Even with these limitations, an internal pilot found that rapid response team activations were reduced by half on the acute care floor and that the average nurse response time decreased, but the system has shown to have a high false-positive rate.[12,13]

4.6 CONCLUSION

We aimed to provide insight in how applying and integrating three types of social or health behavioral information with existing clinical scores can affect the clinical scores for sepsis patients in acute care. Overall, our exploration of integrating SBDH data with existing clinical scoring schemes for potentially sepsis patients in acute care did not yield statistically meaningful influence. We found instances where integrating these more qualitative data types can both increase and decrease performance accuracy. However, it is still important to explore how these factors can influence clinical care and patient outlook as SBDH play a key part in one’s health. However, it may be important to take into consideration other clinical scoring schemes, target a different patient population, like the ED or ICU, or explore other types of social or health behavior data.

4.7 REFERENCES FOR CHAPTER 4

1. Centers for Disease Control and Prevention. Data & Reports | Sepsis | CDC. Accessed August 11, 2019. https://www.cdc.gov/sepsis/datareports/index.html

2. Sepsis. Sepsis - Symptoms and causes - Mayo Clinic. Published November 16, 2018. https://www.mayoclinic.org/diseases-conditions/sepsis/symptoms-causes/syc-20351214

3. O'Brien J. The Cost of Sepsis. The Centers for Disease Control and Prevention Blog. Published November 18, 2016. Accessed February 3, 2020. https://blogs.cdc.gov/safehealthcare/the-cost-of-sepsis/

4. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med*. 1992;20(6):864-874.

5. Balk RA. Systemic inflammatory response syndrome (SIRS): Where did it come from and is it still relevant today? *Virulence*. 2014;5(1):20-26. doi:10.4161/viru.27135

6. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801. doi:10.1001/jama.2016.0287

7. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. *Ann R Coll Surg Engl*. 2006;88(6):571-575. doi:10.1308/003588406X130615

8. Kovach CP, Fletcher GS, Rudd KE, Grant RM, Carlbom DJ. Comparative prognostic accuracy of sepsis scores for hospital mortality in adults with suspected infection in non-ICU and ICU at an academic public hospital. Moreira J, ed. *PLOS ONE*. 2019;14(9):e0222563. doi:10.1371/journal.pone.0222563

9. Roney JK, Whitley BE, Long JD. Implementation of a MEWS-Sepsis screening tool: Transformational outcomes of a nurse-led evidence-based practice project. *Nurs Forum*. 2020;55(2):144-148. doi:10.1111/nuf.12408

10. Tirotta D, Gambacorta M, La Regina M, et al. Evaluation of the threshold value for the modified early warning score (MEWS) in medical septic patients: a secondary analysis of an Italian multicentric prospective cohort (SNOOPII study). *QJM*. Published online January 9, 2017:hcw229. doi:10.1093/qjmed/hcw229

11. Khwannimit B, Bhurayanontachai R, Vattanavanit V. Comparison of the accuracy of three early warning scores with SOFA score for predicting mortality in adult sepsis and septic shock patients admitted to intensive care unit. *Heart & Lung*. 2019;48(3):240-244. doi:10.1016/j.hrtlng.2019.02.005

12. Wolpaw BJ, Kress N, Grant RM, et al. Early Identification and Treatment of Hospital Acquired Sepsis: Success of an EHR-Based System. Presented at the: SGIM 2018.

13. Bean M. This Seattle hospital aims to get sepsis patients on antibiotics within 3 hours — here's how. Becker's Clinical Leadership and Infection Control. Published February 23, 2018. Accessed February 3, 2020. https://www.beckershospitalreview.com/quality/this-seattle-hospital-aims-to-get-sepsis-patients-on-antibiotics-within-3-hours-here-s-how.html

14. Artiga S, Hinton E. *Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity*. KFF: Henry J Kaiser Family Foundation; 2018. Accessed February 7, 2020. http://files.kff.org/attachment/issue-brief-beyond-health-care

15. Social Determinants of Health. HealthyPeople.gov. Accessed February 1, 2020. https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health

16. Social Determinants. Institute for Health Metrics and Evaluation. Accessed February 1, 2020. http://www.healthdata.org/social-determinants

17. Papadopoulou S, Hassapidou M, Katsiki N, et al. Relationships Between Alcohol Consumption, Smoking Status and Food Habits in Greek Adolescents. Vascular Implications for the Future. *Current Vascular Pharmacology*. 2017;15(2):167-173. doi:10.2174/1570161114666161024123357

18. Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. 2015;22(3):553-564. doi:10.1093/jamia/ocu023

19. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19(1):54-60. doi:10.1136/amiajnl-2011-000376

20. Danese MD, Halperin M, Duryea J, Duryea R. The Generalized Data Model for clinical research. *BMC Med Inform Decis Mak*. 2019;19(1):117. doi:10.1186/s12911-019-0837-5

21. FitzHenry F, Resnic FS, Robbins SL, et al. Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. *Appl Clin Inform*. 2015;06(03):536-547. doi:10.4338/ACI-2014-12-CR-0121

22. Papez V, Moinat M, Payralbe S, et al. Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure. *JAMIA Open*. Published online February 4, 2021:ooab001. doi:10.1093/jamiaopen/ooab001

23. Dykes LA, Heintz SJ, Heintz BH, Livorsi DJ, Egge JA, Lund BC. Contrasting qSOFA and SIRS Criteria for Early Sepsis Identification in a Veteran Population. *Fed Pract.* 2019;36(Suppl 2):S21-S24.

24. Lembke K, Parashar S, Simpson S. Sensitivity and Specificity of SIRS, qSOFA and Severe Sepsis for Mortality of Patients Presenting to the Emergency Department With Suspected Infection. *Chest*. 2017;152(4):A401. doi:10.1016/j.chest.2017.08.427

25. For mortality prediction, qSOFA has poor sensitivity, moderate specificity, review finds. Published July 2018. https://acphospitalist.org/weekly/archives/2018/02/07/1.htm

26. Dorsett M, Kroll M, Smith CS, Asaro P, Liang SY, Moy HP. qSOFA Has Poor Sensitivity for Prehospital Identification of Severe Sepsis and Septic Shock. *Prehospital Emergency Care*. 2017;21(4):489-497. doi:10.1080/10903127.2016.1274348

27. Perman SM, Mikkelsen ME, Goyal M, et al. The sensitivity of qSOFA calculated at triage and during emergency department treatment to rapidly identify sepsis patients. *Sci Rep*. 2020;10(1):20395. doi:10.1038/s41598-020-77438-8

28. Machado FR, Assunção MSC de, Cavalcanti AB, Japiassú AM, Azevedo LCP de, Oliveira MC. Getting a consensus: advantages and disadvantages of Sepsis 3 in the context of middle-income settings. *Revista Brasileira de Terapia Intensiva*. 2016;28(4). doi:10.5935/0103-507X.20160068

29. Andaluz D, Ferrer R. SIRS, qSOFA, and organ failure for assessing sepsis at the emergency department. *J Thorac Dis*. 2017;9(6):1459-1462. doi:10.21037/jtd.2017.05.36

30. Faisal M, Scally A, Richardson D, et al. Development and External Validation of an Automated Computer-Aided Risk Score for Predicting Sepsis in Emergency Medical Admissions Using the Patient's First Electronically Recorded Vital Signs and Blood Test Results*: *Critical Care Medicine*. 2018;46(4):612-618. doi:10.1097/CCM.0000000000002967

31. Langley RJ, Tsalik EL, Van Velkinburgh JC, et al. Sepsis: An integrated clinico-metabolomic model improves prediction of death in sepsis. *Science Translational Medicine*. Published online 2013. doi:10.1126/scitranslmed.3005893

32. Sutherland A, Thomas M, Brandon RA, et al. *Development and Validation of a Novel Molecular Biomarker Diagnostic Test for the Early Detection of Sepsis*.; 2011:R149. doi:10.1186/cc10274

33. Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. doi:10.1136/amiajnl-2013-001815

34. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Quality*. Published online 2017. doi:10.1136/bmjoq-2017-000158

Chapter 5. CONCLUSION

5.1 OVERVIEW

Sepsis continues to be a burden for both patients and healthcare institutions as it is a key cause of morbidity, mortality and healthcare costs. Detection and prediction of sepsis with machine learning techniques continue to be a major research area as sepsis remains challenging for providers to detect due to its high prevalence and nonspecific early stage symptoms. Additionally, patient data collection varies across healthcare systems and the changing definitions of sepsis increases variability and complexity.[1–7] A review of the literature indicates that there is currently no consensus amongst the research community on defining a sepsis patient retrospectively, which can directly impact the performance of predictive models as they are dependent on defining outcome variables, or consensus on the best algorithms or detection methods. Although there have been several edits and addendums made to the definition of sepsis over the past few decades, these definitions have been purely based on clinical attributes or lab test results.[8–11] To our knowledge, there has not been a published instance where social and behavioral determinants of health (SBDH) data extracted from unstructured sources were integrated with existing clinical scoring methods to explore the impact and influence of these additional non-clinical data. However, recently a protocol proposal has been published aiming to explore how social determinants of health are associated with the development of sepsis in adults.[12]

Concurrently, in recent years, collection of SBDH data has increased making these data more complete for research, whether through structured sources like International Classification of Disease (ICD) Z-codes and intake forms, or unstructured sources like physician notes.[13,14(p10),15] The need to analyze these more demographic and qualitative data are becoming more present as SBDH are important influences on overall health.[16,17] Additionally, recent research has identified various SBDH, including housing, income, tobacco and alcohol use, race, and education, to have an impact on sepsis prevalence and mortality rate among more vulnerable populations, which include the uninsured, homeless, and those with multiple comorbidties.[18–23] The papers presented in this dissertation addressed the identified gaps in research by exploring the impact of integrating SBDH data with established clinical scoring methods on acute care patients who present sepsis-like qualities. Specifically, we explored the impact of generating housing stability, alcohol use, and tobacco use scores with qSOFA, SIRS, and MEWS scoring methods.

The first paper resulting from this work (Chapter 2) presented findings from a literature review which aimed to identify and aggregate the status quo regarding sepsis detection methods and sepsis definitions. In this work, we systematically reviewed and detailed the current literature to identify relevant studies that used a predictive analytics solution to predict the onset or mortality of sepsis in hospitalized patients. We identified a total of 31 studies and detailed the methods and models that each study utilized and also described their main findings. As each study presented results, identified patient populations, and defined sepsis onset or mortality differently, it was important to note the challenges with comparing results between studies. Overall, we found wide variation in the domain of predictive analytics tools for sepsis patients, from feature and population size to selection of prediction method or analytic algorithm. Even though there were discrepancies between methods and there were no clear best practices to extract from the review, it was evident that implementing predictive analytics tools for sepsis detection and prediction for onset or mortality were still beneficial in the battle of sepsis.[24]

The second and third paper (Chapters 3 and 4) from the dissertation presented findings from the dissertation study. SBDH data can be collected through various avenues. However, it has been noted in the literature that purely relying on structured data, typically diagnosis codes, may not be complete or entirely representative of the patient population.[13] Therefore, we decided to explore extracting these SBDH data features from unstructured data, most notably clinical notes from admission and the emergency department. We found that these notes at Harborview Medical Center were prime for extracting three concepts, housing stability, alcohol use, and tobacco use, which corresponded well with the King County Community Health Needs Assessment (CHNA) report.[25] After manually sifting through 25 notes, we discovered that these selected clinical notes were written in a fairly structured manner which made the process of extracting the text feasible. We identified text classification methods that were commonly used in the data science field and applied them to these clinical notes. Additionally, we aimed to use more simple text classification techniques as more complex methods would be harder to interpret in the medical space. Overall, we found that text classification worked well with these selected clinical notes and identified features of interest. Furthermore, we compared the unstructured data extraction results with extracting the same concepts through structured data. We found that for housing stability and alcohol use, classification from clinical notes provided more data than structured data. However, tobacco use data was more complete with structured data as tobacco use screening is a core CMS measure and a part of Meaningful Use; alcohol dependence is also a part of this measure, but general alcohol use screening is not.[26] Altogether, our simplified approach has shown that open source simple text classifiers can be used to predict text sentiment for SBDH and can supplement current structured sources to provide a more complete social history for patients, especially when SBDH structured data are not complete or regularly recorded.

The third paper focused specifically on integrating the results from Chapter 2 with sepsis scoring methods commonly used in the acute care setting. We aimed to provide novel insight in how applying and integrating three types of SBDH information with existing clinical scores can affect the clinical scores for sepsis patients in acute care. With baselines scores, we found that SIRS had a higher accuracy than qSOFA and MEWS for identifying potential sepsis patients in acute care. But, once we integrated the SBDH scores that were created in Chapter 3, we found that qSOFA has a slightly higher accuracy compared to SIRS and MEWS. However, in order to reduce missing data of necessary features for each scoring method, patient encounters that did not have values for a required feature per scoring method were excluded; therefore, a different subset of the initially identified population were used for each calculation making it harder to directly compare performance across the scoring methods. Overall, we found that the integration of SBDH data from unstructured clinical notes did not improve accuracy or performance in both a longitudinal approach or prior year approach. In fact, the use of housing stability, tobacco use, or alcohol use slightly lowered performance overall when integrated with existing clinical scoring schemes and using purely SBDH did not correlate well with our definition of likely having sepsis in the acute care setting. We found instances where integrating these more qualitative data types can both increase and decrease performance accuracy. Therefore, it is still important to explore how these factors can influence clinical care and patient outlook as social determinants play a key part in one’s health.

5.2 LIMITATIONS

All research has limitations within the methods. Here, we highlight the limitations of each of the three aims.

5.2.1.1 Aim 1 Limitations

The systematic scoping review query was designed to be broad in order to capture the variety of methods and predictive analytic solutions being used and created for detecting sepsis onset or mortality. Furthermore, MesH terms, which catalog and index the articles in PubMed, were not used in the query. Utilizing MesH terms could have potentially narrowed down further our initial selection of articles. Additionally, if we narrowed the objective of our review to solely include studies that aimed to decrease sepsis mortality or onset, the search would have brought in different studies and examples. Furthermore, the definition of sepsis is ever changing, and this solely marks, generally, what had been done up until the search date. PubMed is not the only database for research articles. Our review only queried PubMed; searching in other literature databases such as Google Scholar, Excerpta Medica Database (EMBASE), Medical Literature Analysis and Retrieval System Online (MEDLINE), or the Cumulative Index to Nursing and Allied Health Literature (CINAHL) could have yielded a different set of results potentially enriching our findings. Collaborating with a university librarian could have also made the search more robust.

5.2.1.2 Aim 2 Limitations

There were two main points in our process that required manual attention: (1) EHR review and (2) labelling of features within the clinical text. Manual EHR review was done to ensure the completeness of social history information and to confirm that these quantitative data were in a consistent note location before implementing a widespread text extraction. Initially, we validated the completeness of the data with a random set of ten patients, reviewing social history sections of both admission and emergency department notes, but later expanded our validation check to a total of 25 patients. We performed a statistical calculation and determined that the presence of these data features in notes within a random set 25 patient indicated a high level of confidence of data completeness. Furthermore, labelling features manually was a time-consuming process and was performed only by one person. Having multiple team members would provide a more accurate and possibly more consistent labelling of features to train the model for text classification as there would be consensus between labelers.

Although we aimed to create generalizable workflow, it was limited due to unique nuances in note-taking language, such as specific shelter names, addresses, or abbreviations that might not be standardized in the medical field. Furthermore, because patients can downplay or withhold information about the social challenges they face, incoming data streams can inhibit the potential of text classification. Our method also depended on the patient having been seen within the healthcare system at some point, inpatient or outpatient, within the past five years. Additionally, we did not incorporate outside ontologies, such as UMLS or MetaMap, as we were interested in creating a simple text classification approach that did not need to rely on outside entities.

5.2.1.3 Aim 3 Limitations

From Chapter 2 (Aim 1), we found that there were many ways to define and confirm a sepsis diagnosis. Given our data extract, we decided to create our confirmation definition that utilized a combination of expert data extracted from a nurse checklist, diagnosis codes, and the presence of antibiotics. Therefore, our confirmation definition of a sepsis diagnosis did not aggregate severity levels definitions, including septic shock and severe sepsis. Because we were interested in exploring the impact of SBDH data with clinical scoring methods commonly used for sepsis, we determined that the we did not need to explore the varying levels of severity of sepsis as we wanted to explore the general influence and capabilities of the existing scoring systems. Additionally, we were not calculating clinical scores as an early-warning but rather taking the highest calculated score during an encounter and comparing it directly with our confirmation definition. This minimized bias from potentially overexposed patients, such as those with longer stays or those who had more frequent vital signs or laboratory results. Additionally, patients with missing data or lack of complete data were removed potentially reducing our patient population.

Another slight limitation was that ICD codes were not always numerically coded in the database under the diagnosis table; however, were correctly coded in the billing table. Therefore, we had to approach the diagnosis table via a string search. This search could have potentially missed instances that did not explicitly mention sepsis, septicemia, or urosepsis in the diagnosis description string.

Identifying our patient population differed slightly from other existing literature. A number of past studies included patients based off of their location at a specific time and then filtered and excluded based on their defined criteria. Our process was reversed; we queried and included patients based on their vital signs and lab results first, and then filtered patients by location. Furthermore, our results from Chapter 3 (Aim 2) were calculated independently from the query used in Chapter 4 (Aim 3) and were merged via patient identifier, data, and hour. This can potentially lead to data that is available to be not merged as vital signs and laboratory results may not occur at the same hour as the intake note. In any situation like this, the social features were not merged as there was no equivalent hour to merge on.

In an effort to include expert data, we extracted data from a nursing checklist was designed to capture both binary and qualitative data from nurses who suspected an infection. Due to the complex nature of notes, these data were not utilized but could potentially provide another layer of expert data. However, alerts for this system are only generated after manual entry of lab results or vital signs as the alert system is not directly integrated with the EHR, creating a slight delay.

5.3 FUTURE WORK AND RECOMMENDATIONS

5.3.1.1 Immediate Considerations

Future work can build upon the established workflow and results presented in this dissertation. To better understand the state of sepsis predictive analytics, a new search query could be performed as new research in the past few years could have evolved. Furthermore, searching across multiple databases would potentially provide more comprehensive and complete results. Our search query erred towards maximizing recall as we did not want to miss or exclude any studies; however, future searches may want to err towards precision to identify specific articles and results. Collaborating with a librarian and having multiple team members analyzing titles and abstracts would provide better insight and more robust search queries or results. Utilizing software tools, such as Rayyan or Covidence, would provide structured, easy-to-use, and streamlined collaboration.[27]

Community needs are constantly changing as the health of the community is not static. Currently, the King County CHNA has identified obesity, healthcare access, insurance status and drug use as other potential SBDH information to explore. These data types would be stored in different areas of the EHR and within different notes. It would be interesting to see if our designed workflow presented in Chapter 3 could be applicable and generalized to meet the needs of other SBDH data. Although we aimed to create a simplified framework to extract SBDH data from clinical notes, more complex methods such as convoluted neural networks and more advanced NLP part of speech tagging may be worth exploring as they may help improve accuracy and precision of the classification. As more notes become available for patients, it will also be important to keep in mind the potential bias of having more notes present from sicker patients and evaluating ways to reduce this bias.

Furthermore, as UW Medicine has just recently transitioned all EHR systems to Epic, it would be of value to explore how the shift in systems impact the established text extraction and classification workflow. Similarly, testing the workflow at another healthcare institution would show the generalizability and implantability of the described workflow. With the adoption of the OMOP common data model, we expect the workflow to work and transition seamlessly in theory, but practice may prove otherwise. Furthermore, as more SBDH continue to be collected from intake forms, patient portals, and providers, it may be important to identify how to combine both qualitative and quantitative data and determine which data sources are more trustworthy and assign weights to create a holistic scoring system.

Integrating SBDH data is still important to explore. Although our initial results presented in this dissertation showed that our selected social features did not yield statistically meaningful results, other SBDH could result differently. Furthermore, the integration of social determinants and clinical data may prove to be more influential in a different hospital setting. Collaborating with experts in the field may also improve the presented work. Although we did not have expert review of EHR data, we tried to supplement this shortcoming by integrating qualitative data from a nursing checklist, essentially adding expert data. However, having a few experts review the EHR and collectively identify and classify notes as well as scoring methods would substantially increase the validity of the results. As sepsis research evolves, it would be important to take future definitions and features of importance into consideration. Recent literature has shown hydrocortisone, vitamin C, and thiamine as newer sepsis treatment workflows, although may not provide more rapid resolution of septic shock, may still be worth exploring and provide additional insight with other non-clinical data.[28,29]

5.3.1.2 Long-Term Directions

Current research has also indicated that incorporating individual level or community level SBDH features may not impact patient length of stay and readmission rates for the general patient population.[22,29] However, it was shown that specific vulnerable subpopulations from the general patient population benefited from the incorporation of SBDH data, including Medicaid patients, obese patients, and patients aged 65 and older.[29] Therefore, with our work, it similarly may be worth exploring these vulnerable subpopulations rather than the general patient population in the acute care setting. Targeted subpopulations could also include those with two or more comorbidities, the uninsured, or immunocompromised patients.[31–33] These more vulnerable populations are less likely to seek consistent preventative care as their priorities are different, services which can accumulate requiring more invasive treatment later in life. It may also be interesting to explore how to flag or score the frequency of visits by patient to extract those who do not consistently seek preventative care. Therefore, it is important to explore how the integration of SBDH data can impact not only sepsis prediction for these vulnerable populations, but also readmission and post-acute care mortality risk.

We sourced data from solely one medical center. Patients might have had encounters or other visit types in neighboring hospitals and healthcare systems in the region. The lack of data sharing between institutions prevents holistic collection of SBDH data. Data completeness is vitally important to the quality and accuracy of models that are dependent on big data. Poor data quality and completeness lead to lower utilization and the lack of data can potentially lead to mistakes in the decision-making process; additionally, since there is no single or standardized source for SBDH data, the diversity of data and complexity of the associated data structures increase the difficulty and bottlenecks for data integration.[34] The lack of a standardized methodology to collect and store all SBDH data will limit the potential of this research field. Additionally, SBDH factors are constantly changing for patients as their behaviors can change depending on their circumstance. Being able to aggregate these data and create adaptable models is crucial as these features are never static. Furthermore, public health, outreach services fluctuate, and even medical terminology can over time. Creating a method or utilizing an API to update the list of community shelters and other places for homeless services would be necessary to maintain an accurate understanding of a patients housing status. A data dictionary mapping older medical terminology and shorthand with current language can also aid with better extracting of clinical notes.

Our analysis follows a heuristic that hypothesizes a correlation between SBDH and the likelihood for an acute care patient to develop sepsis. However, it may also be worth exploring the inverse where one identifies a list of features (vital signs, lab results, and SBDH factors) and apply machine learning techniques to tease out the weights and importance for each feature to identify which are more influential in determining the development of sepsis in acute care. However, this approach would be limited to structured data sources only, which we found to be incomplete for two out of the three SBDH that were selected in our analysis. Utilizing clinical text to extract and classify SBDH sentiment would not be feasible from this workflow but may still be worth exploring in the future as SBDH data collection continues to improve.

5.4 IMPLICATIONS

5.4.1.1 Implications for Healthcare Institutions

Although our approach integrating SBDH with existing sepsis definitions and methods did not yield an improvement of sepsis prediction, it is still important to continue to explore how SBDH can affect patient health outcomes. Past research has consistently pointed towards SBDH impacting patient health and outcomes. However, collection of SBDH can be a major limiting factor in the ability to model and integrate these data. There has not been a standardized collection process for SBDH data across the institution, whether it is recorded through notes or electronic forms. Additionally, many times, SBDH data may not be asked due to patient condition or it might not be updated regularly.

Providers and healthcare institutions should strive to collect SBDH data more regularly even if the data fields are not empty as SBDH status can change. These intake procedures should be present and not optional; currently, only language preference must be completed due to translation and interpreter laws in place. Additionally, educating patients to utilize patient portals and update information via these portals can provide more current SBDH information. However, we should note that vulnerable populations would most likely not be the primary audience to utilize this feature, and this is the subpopulation that arguably needs more attention.

5.4.1.2 Implications for Public/Population Health and Policy

Patients can seek care from multiple providers. Data interoperability between health systems has historically been challenging on multiple levels, including data structures and semantics. Creating policies to ensure reporting standards could improve data interoperability. Having an integrated data sharing option for SBDH data can help eliminate data completeness concerns. Creating strategies to coordinate clinical and social services can aid community-level efforts to address SBDH concerns. Improving public health outreach and initiatives for these vulnerable populations can potentially reduce health disparities and potentially reduce the burden of sepsis and other diseases. Additionally, COVID-19 has shown the importance of being able to identify and provide support to vulnerable populations.

5.5 CONCLUSIONS

Overall, we found that there are currently, to our knowledge, standard or uniform method for predicting sepsis onset or mortality within the hospital setting, whether it be in the intensive care unit or emergency department. Furthermore, we found that the existing literature solely focus on clinical features, such as vital signs and laboratory results. Recent literature has suggested that SBDH can influence overall health. Therefore, we explored how these non-clinical data can impact clinical scoring methods commonly used on sepsis patients in acute care. Although we found success in extracting SBDH data, creating a score and integrating it with existing clinical scoring methods for sepsis patients did not yield a significant difference. While this stands in contrast to existing research suggesting the importance of SBDH and clinical outcomes, it seems for this selected disease, there may be other factors more worth investigating.

5.6 REFERENCES FOR CHAPTER 5

1. Usman OA, Usman AA, Ward MA. Comparison of SIRS, qSOFA, and NEWS for the early identification of sepsis in the Emergency Department. *The American Journal of Emergency Medicine*. 2019;37(8):1490-1497. doi:10.1016/j.ajem.2018.10.058

2. Bhattacharjee P, Edelson DP, Churpek MM. Identifying Patients With Sepsis on the Hospital Wards. *Chest*. 2017;151(4):898-907. doi:10.1016/j.chest.2016.06.020

3. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock. *Critical Care Medicine*. 2013;41(2):580-637. doi:10.1097/CCM.0b013e31827e83af

4. Angus DC, van der Poll T. Severe Sepsis and Septic Shock. *N Engl J Med*. 2013;369(9):840-851. doi:10.1056/NEJMra1208623

5. Levy MM, Artigas A, Phillips GS, et al. Outcomes of the Surviving Sepsis Campaign in intensive care units in the USA and Europe: a prospective cohort study. *The Lancet Infectious Diseases*. 2012;12(12):919-924. doi:10.1016/S1473-3099(12)70239-6

6. Moore JX, Donnelly JP, Griffin R, et al. Black-white racial disparities in sepsis: a prospective analysis of the REasons for Geographic And Racial Differences in Stroke (REGARDS) cohort. *Critical Care*. 2015;19(1). doi:10.1186/s13054-015-0992-8

7. Amland RC, Hahn-Cover KE. Republished: Clinical Decision Support for Early Recognition of Sepsis. *Am J Med Qual*. 2019;34(5):494-501. doi:10.1177/1062860619873225

8. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med*. 2003;29(4):530-538. doi:10.1007/s00134-003-1662-x

9. Vincent J-L, Opal SM, Marshall JC, Tracey KJ. Sepsis definitions: time for change. *The Lancet*. 2013;381(9868):774-775. doi:10.1016/S0140-6736(12)61815-7

10. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801. doi:10.1001/jama.2016.0287

11. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med*. 1992;20(6):864-874.

12. Machon C, Sheikh F, Fox-Robichaud A. Social determinants of health associated with the development of sepsis in adults: a scoping review protocol. *BMJ Open*. 2020;10(10):e039146. doi:10.1136/bmjopen-2020-039146

13. Torres JM, Lawlor J, Colvin JD, et al. ICD Social Codes: An Underutilized Resource for Tracking Social Needs. *Medical Care*. 2017;55(9):810-816. doi:10.1097/MLR.0000000000000764

14. Truong HP, Luke AA, Hammond G, Wadhera RK, Reidhead M, Joynt Maddox KE. Utilization of Social Determinants of Health ICD-10 Z-Codes Among Hospitalized Patients in the United States, 2016–2017. *Medical Care*. 2020;58(12):1037-1043. doi:10.1097/MLR.0000000000001418

15. Clark MA, Gurewich D. Integrating Measures of Social Determinants of Health Into Health Care Encounters: Opportunities and Challenges. *Medical Care*. 2017;55(9):807-809. doi:10.1097/MLR.0000000000000788

16. Braveman P, Gottlieb L. The Social Determinants of Health: It's Time to Consider the Causes of the Causes. *Public Health Rep*. 2014;129(1_suppl2):19-31. doi:10.1177/00333549141291S206

17. Daniel H, Bornstein SS, Kane GC, for the Health and Public Policy Committee of the American College of Physicians. Addressing Social Determinants to Improve Patient Care and Promote Health Equity: An American College of Physicians Position Paper. *Ann Intern Med*. 2018;168(8):577. doi:10.7326/M17-2441

18. Alroumi F, Abdul Azim A, Kergo R, Lei Y, Dargin J. The impact of smoking on patient outcomes in severe sepsis and septic shock. *j intensive care*. 2018;6(1):42. doi:10.1186/s40560-018-0312-x

19. Bagaitkar J, Demuth DR, Scott DA. Tobacco use increases susceptibility to bacterial infection. *Tob Induced Dis*. 2008;4(1):12. doi:10.1186/1617-9625-4-12

20. Huttunen R, Laine J, Lumio J, Vuento R, Syrjänen J. Obesity and smoking are factors associated with poor prognosis in patients with bacteraemia. *BMC Infect Dis*. 2007;7(1):13. doi:10.1186/1471-2334-7-13

21. Gustot T, Fernandez J, Szabo G, et al. Sepsis in alcohol-related liver disease. *Journal of Hepatology*. 2017;67(5):1031-1050. doi:10.1016/j.jhep.2017.06.013

22. Schnegelsberg A, Mackenhauer J, Nibro HL, Dreyer P, Koch K, Kirkegaard H. Impact of socioeconomic status on mortality and unplanned readmission in septic intensive care unit patients. *Acta Anaesthesiol Scand*. 2016;60(4):465-475. doi:10.1111/aas.12644

23. Kempker JA, Kramer MR, Waller LA, Martin GS. Risk Factors for Septicemia Deaths and Disparities in a Longitudinal US Cohort. *Open Forum Infect Dis*. 2018;5(12):ofy305. doi:10.1093/ofid/ofy305

24. Teng AK, Wilcox AB. A Review of Predictive Analytics Solutions for Sepsis Patients. *Appl Clin Inform*. 2020;11(03):387-398. doi:10.1055/s-0040-1710525

25. King County Community Health Needs Assessment 2018/2019. https://www.kingcounty.gov/depts/health/data/community-health-indicators/~/media/depts/health/data/documents/2018-2019-Joint-CHNA-Report.ashx

26. All eCQMs associated with: Prevention and Treatment of Opioid and Substance Use Disorders. https://ecqi.healthit.gov/mcw/list/meaningful-measure/prevention-and-treatment-opioid-and-substance-use-disorders

27. Cleo G, Scott AM, Islam F, Julien B, Beller E. Usability and acceptability of four systematic review automation software packages: a mixed method design. *Syst Rev*. 2019;8(1):145. doi:10.1186/s13643-019-1069-6

28. Walter KL, Seymour CW. Hydrocortisone, Vitamin C, and Thiamine for Treatment of Sepsis: Making Evidence Matter. *JAMA*. 2021;325(8):730. doi:10.1001/jama.2020.26029

29. Fujii T, Luethi N, Young PJ, et al. Effect of Vitamin C, Hydrocortisone, and Thiamine vs Hydrocortisone Alone on Time Alive and Free of Vasopressor Support Among Patients With Septic Shock: The VITAMINS Randomized Clinical Trial. *JAMA*. 2020;323(5):423. doi:10.1001/jama.2019.22176

30. Zhang Y, Zhang Y, Sholle E, et al. Assessing the impact of social determinants of health on predictive models for potentially avoidable 30-day readmission or death. *PLoS One*. 2020;15(6):e0235064. doi:10.1371/journal.pone.0235064

31. Japiassú AM, Amâncio RT, Mesquita EC, et al. Sepsis is a major determinant of outcome in critically ill HIV/AIDS patients. *Crit Care*. 2010;14(4):R152. doi:10.1186/cc9221

32. Kumar G, Taneja A, Majumdar T, Jacobs ER, Whittle J, Nanchal R. The Association of Lacking Insurance With Outcomes of Severe Sepsis: Retrospective Analysis of an Administrative Database*. *Critical Care Medicine*. 2014;42(3):583-591. doi:10.1097/01.ccm.0000435667.15070.9c

33. Wang HE, Shapiro NI, Griffin R, Safford MM, Judd S, Howard G. Chronic Medical Conditions and Risk of Sepsis. Gold JA, ed. *PLoS ONE*. 2012;7(10):e48307. doi:10.1371/journal.pone.0048307

34. Cai L, Zhu Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *CODATA*. 2015;14(0):2. doi:10.5334/dsj-2015-002

APPENDIX

Table A4.1: List of antibiotics

List of antibiotics used for patients if sepsis is suspected
imipenem-cilastatin
meropenem
cefotaxime
ceftazidime
ceftriaxone
cefepime
ceftaroline
moxifloxacin
levofloxacin
amoxicillin-clavulanate
ampicillin-sulbactam
piperacillin-tazobactam
ticarcillin-clavulanate
amikacin
gentamicin
tobramycin
aztreonam
cefazolin
cefuroxime
ciprofloxacin
clindamycin
daptomycin
vancomycin
linezolid
azithromycin
erythromycin
ampicillin
nafcillin
penicillin g

Table A4.2: Aim 3 prediction distribution (confusion matrix), longitudinal method

qSOFA w/ SBDH	TN = 29,971 FP = 39,784 FN = 403 TP = 3,216
SIRS w/ SBDH	TN = 25,214 FP = 38,313 FN = 66 TP = 2,060
MEWS w/ SBDH	TN = 9,102 FP = 15,613 FN = 400 TP = 2,447