Predictive Modeling for Opioid-Related Overdose Prevention: Towards Safer Opioid Prescribing and Guiding Treatment Selection for Opioid Use Disorder

By

Leigh Anne Tang

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

In

Biomedical Informatics

December 14, 2024

Nashville, Tennessee

Approved:

Colin G. Walsh, MD, MA, FAMIA, FACMI, FIAHSI

Jessica S. Ancker, PhD, MPH, FACMI

Lisa Bastarache, MS

Qingxia (Cindy) Chen, PhD

Kristopher A. Kast, MD

Copyright © 2024 Leigh Anne Tang
All Rights Reserved

To Tom, Bing, and Sandy

ACKNOWLEDGMENTS

I am grateful for the support of my committee members: Dr. Colin Walsh, Dr. Jessica Ancker, Lisa Bastarache, Dr. Cindy Chen, and Dr. Kristopher Kast. I am incredibly thankful to Drs. Colin Walsh and Jessica Ancker, who have mentored and guided me throughout my training. It has been such a pleasure to work with and learn from Lisa Bastarache since the beginning of my time at the Vanderbilt Department of Biomedical Informatics (DBMI). I am grateful to Dr. Cindy Chen, who has never hesitated to advise me on both side projects and dissertation work, and Dr. Kristopher Kast for his willingness to explain the nuances of clinical care to me time and time again.

Several people supported the work in this dissertation. A significant portion of this work would not have been possible without Drs. Sutapa Mukhopadhyay and Jessica Korona from the Tennessee Department of Health (TDH), who helped me understand public health surveillance and entrusted me with TDH resources. Michael Ripperger was always happy to answer any questions I had related to the Walsh Lab. Dr. Tom Reese, Michelle Gomez, Uday Suresh, and Alex Becker contributed to various aspects of the qualitative study included in this dissertation.

Many people outside of my committee have offered invaluable support and wisdom. Drs. Lina Sulieman, Laurie Novak, Megan Salwei, and Joshua Smith have been incredibly generous with their time, insight, and encouragement. I would like to thank Nick Hyer in the Vanderbilt University Graduate and Postdoc Academic Success (GPAS) for organizing programs like Write On! to support trainees like me. The DBMI training program would not be the same without Dr. Kim Unertl and Rischelle Jenkins, who have always made themselves available and have never hesitated to advocate on my behalf. Lastly, I am so fortunate to have the support of friends and

family. Special thanks to Alex Becker for always making time for me and to my husband Tom for reminding me to have fun (and eat) throughout this journey.

This work was supported by the National Library of Medicine (T15LM007450), Vanderbilt University Graduate Leadership Institute Dissertation Enhancement Grant, and Vanderbilt University Medical Center Randolph Miller Award.

TABLE OF CONTENTS

ACKNOWLEDGMENTS ..... **iv**

LIST OF TABLES ..... **viii**

LIST OF FIGURES ..... **ix**

1 INTRODUCTION AND OVERVIEW **1**
1.1 Specific Aims 2
1.2 Dissertation Roadmap 5
2 BACKGROUND **7**
2.1 A Primer on Opioids 7
2.2 The Ongoing Opioid Overdose Epidemic 8
2.3 Evidence-Based Interventions for Overdose Prevention 9
2.4 Towards the Safer Prescribing of Opioids 11
2.4.1 High-Risk Opioid Prescribing 11
2.4.2 Overdose Surveillance Systems 12
2.4.3 Predictive Analytics Using Prescription Drug Monitoring Program Data 13
2.5 Towards Guiding Treatment Selection for Opioid Use Disorder 13
2.5.1 Treating Opioid Use Disorder with Medication 14
2.5.2 Initiating Medication for OUD in the Hospital 15
2.5.3 Treatment Selection Models for More Personalized Care in Mental Health 16
2.5.4 Challenges of Implementing Predictive Models as Clinical Decision Support 18
2.6 Dissertation Overview 19
3 PROSPECTIVE VALIDATION OF ENSEMBLE LEARNING MODELS FOR PREDICTING OPIOID-RELATED OVERDOSE USING TENNESSEE STATEWIDE DATA **20**
3.1 Introduction 20
3.2 Methods 22
3.2.1 Data Sources 23
3.2.2 Outcomes 24
3.2.3 Model Predictors 24
3.2.4 Model Recalibration 26
3.2.5 Assessment of Model Performance at the Prescription Level 26
3.2.6 Aggregation of Prescription-Level Predictions to the County Level 27
3.2.7 Comparison of Aggregated County-Level Predictions to Observed Overdose Rates 27
3.3 Results 28
3.3.1 Cohort 28
3.3.2 Outcomes 28
3.3.3 Model Performance Assessment 30
3.3.4 Risk Concentration Analysis 30
3.3.5 Subgroup Analysis 30
3.3.6 Comparison of County-Level Predictions to Observed Overdose Rates 33
3.4 Discussion 35
3.4.1 Models Performed Similarly on Newer Data 36
3.4.2 Models Needed Recalibrating 36
3.4.3 Identifying High-Risk Prescriptions from Model Predictions 37
3.4.4 Performance Varied Across Subgroups 38
3.4.5 Limitations 39
3.4.6 Future Work 39
3.5 Conclusion 40
4 A TREATMENT SELECTION MODEL FOR OPIOID USE DISORDER USING ELECTRONIC HEALTH RECORD AND ZIP-LEVEL DATA **41**
4.1 Introduction 41
4.2 Methods 43
4.2.1 Data Sources 43
4.2.2 Cohort Definition 44
4.2.3 Outcomes 44
4.2.4 Model Predictors 45

4.2.5 Univariate Statistical Analysis ... 47
4.2.6 Modeling Plan ... 48
4.2.7 Model Performance Assessment ... 49
4.2.8 Predicting Treatment Response for each MOUD ... 50
4.2.9 Post-Hoc Analysis ... 51
4.2.10 Experimental Setup ... 52
4.3 Results ... 52
4.3.1 Comparison of Individuals on Methadone vs. Buprenorphine ... 53
4.3.2 Cohort Outcomes ... 56
4.3.3 Model Performance ... 57
4.3.4 Top Predictors of MOUD Response ... 59
4.3.5 Probabilities of MOUD Response ... 59
4.4 Discussion ... 61
4.4.1 Social Determinants of Health Not Well-Represented ... 62
4.4.2 Top Predictors Suggested Clinical Complexity ... 63
4.4.3 Local Prescribing Practices Influenced Training Data ... 64
4.4.4 Limitations ... 64
4.5 Conclusion ... 65
5 CLINICIAN PERSPECTIVES ON A PREDICTIVE MODEL FOR RECOMMENDING OPIOID USE DISORDER TREATMENT ... 66
5.1 Introduction ... 66
5.2 Methods ... 68
5.2.1 Setting ... 69
5.2.2 A Predictive Model for MOUD Selection ... 69
5.2.3 Data Collection ... 70
5.2.4 Data Analysis ... 71
5.3 Results ... 72
5.3.1 Socio-Technical Model Dimensions ... 72
5.3.1.1. People ... 73
5.3.1.2. External Rules, Regulations, and Pressures ... 75
5.3.1.3. Hardware and Software ... 77
5.3.1.4. Clinical Content ... 78
5.3.1.5. System Measurement and Monitoring ... 79
5.3.1.6. Human Computer Interface ... 80
5.3.1.7. Internal Organizational Policies and Procedures ... 82
5.3.1.8. Workflow and Communication ... 83
5.4 Discussion ... 87
5.4.1 Non-Specialists Trusted Specialists and the Institution ... 87
5.4.2 Clinicians Expected the Predictive Model to Think Like a Clinician ... 88
5.4.3 Clinical Reasoning for MOUD Decision Requires Considering Information Not Well-Captured by the EHR ... 88
5.4.4 Predictive Models for OUD Treatment Should Be Designed to Foster Appropriate Trust ... 89
5.4.5 Limitations ... 90
5.5 Conclusion ... 90
6 CONCLUSION ... 92
6.1 Innovation ... 92
6.2 Implications for the Domain of Opioids ... 94
6.3 Implications for the Field of Biomedical Informatics ... 96
6.4 Conclusion ... 97
APPENDICES ... 98
A Predictors for Overdose Risk Prediction Models ... 98
B Modeling Plan and Additional Analyses for MOUD Treatment Selection Model ... 99
C Conceptualization and Expansion of Sittig and Singh Socio-Technical Model ... 103
REFERENCES ... 105

LIST OF TABLES

Table Page

1. Prospective validation results for ridge regression ensembled models. 29
2. Risk concentration of ensembled fatal and nonfatal model predictions in test set. 31
3. Model metrics for overdose risk prediction models, by subgroup..... 31
4. Univariate comparisons of individuals on methadone vs. buprenorphine. 54
5. Adverse outcomes for entire cohort and individuals started on each MOUD. 56
6. Median and bootstrapped confidence intervals for model performance metrics in the test set. 57
7. Layout of semi-structured interview guide. 70
8. Calibration metrics for MOUD treatment selection models using EHR data only..... 101
9. Expanded definitions for 3 dimensions in the Sittig and Singh socio-technical model.. 104

LIST OF FIGURES

Figure Page

1. Overdose risk prediction model recalibration and prediction aggregation steps. 25
2. Fatal Overdose Model – Comparison of predicted high-risk prescriptions vs. observed fatal overdose rates in 2022. 34
3. Nonfatal Overdose Model - Comparison of predicted high-risk prescriptions vs. observed nonfatal overdose rates in 2022. 35
4. Treatment selection model architectures adapted from Sheu et al. to include early and late fusion variations. 49
5. Derivation of differential MOUD response probability. 50
6. Individuals initiated on MOUD in the hospital at VUMC from 2018-2023. 53
7. Global SHAP scores and local SHAP values for top 15 predictors in the deep learning model. 60
8. Differential MOUD response probabilities for individuals initiated on each MOUD, specialists only. 61
9. Sittig and Singh socio-technical model dimensions. 71
10. Predictors at the prescription, ZIP, and county levels for the fatal and nonfatal opioid-related overdose risk prediction models. 98
11. Model and data configurations for the different MOUD treatment selection models tested. 99
12. Matthew's Correlation Coefficient formula. 100
13. Receiver operating characteristic curve for the deep learning model with EHR data only. 100
14. Precision-recall curve for the deep learning model with EHR data only. 101
15. Differential MOUD response probabilities for individuals initiated on each MOUD, all clinicians in test set. 102
16. MOUD response probabilities and top predictors for an individual who was prescribed buprenorphine. 102
17. Overlapping dimensions of the Sittig and Singh socio-technical model. 103

CHAPTER 1

INTRODUCTION AND OVERVIEW

In 2022 alone, opioid-related overdoses resulted in 81,806 deaths in the United States, the most recorded in any year[1]. The opioid epidemic was initially driven by prescription opioids in 1999 but has since seen increases in overdoses linked to synthetic opioids and stimulants[1]. The continued toll of the opioid epidemic calls for increased prevention efforts. Together, researchers and community leaders have developed the Opioid-Overdose Reduction Continuum of Care Approach (ORCCA), which details several evidence-based practices for reducing opioid-related overdose[2]. ORCCA outlines interventions that span the public health and individual levels (from increased naloxone distribution to improved delivery of medication for opioid use disorder), suggesting that the opioid epidemic calls for a combination of different interventions. As such, in my dissertation, we aimed to inform overdose prevention through two interrelated projects, one at the public health level and the other at the individual level.

At the public health level, one state that has been heavily impacted by the epidemic is Tennessee. In 2022, Tennessee recorded the third highest overdose death rate in the United States[3]. The prevalence of opioid-related overdose deaths highlights the need for improved prevention and intervention. Prevention traditionally has been guided by observational studies of statewide opioid prescribing patterns, but overdose surveillance data can often be delayed[4]. Meanwhile, prescription drug monitoring program (PDMP) data offer near real-time information. Predictive models using PDMP data have the potential to complement existing retrospective studies to guide more timely prevention. States like Maryland, Oregon, and Pennsylvania have developed overdose prediction models using statewide data[5–8]. Ripperger et al. were the first to

use data from a southern state for overdose risk prediction[9]. However, little is known about the performance of statewide overdose risk prediction models on data from nearly 10 years later.

At the patient level, individuals with opioid use disorder (OUD) are at an increased risk for opioid-related overdose. The first line of treatment for OUD is medication, yet only 25% of individuals with OUD reported receiving medication for OUD (MOUD)[10]. Typically, either methadone or buprenorphine is prescribed as MOUD. Studies have reported a wide range of MOUD retention rates, suggesting that treating OUD remains a challenge[11]. The decision between methadone and buprenorphine is complicated by factors such as insurance coverage, access to methadone clinics, and patient preference. Additionally, lower rates of MOUD initiation have been observed when specialists like addiction psychiatrists have not been involved in OUD care decisions[12,13]. A predictive model recommending MOUD could help empower non-specialists to make decisions without needing to rely on specialists. Previous predictive modeling studies have focused on MOUD discontinuation or only one MOUD[14–16]. No study to our knowledge has attempted to inform treatment selection for OUD. A model predicting treatment response could increase the rates of MOUD initiation and ultimately help reduce overdose rates.

1.1 Specific Aims

This work studied the development, validation, and user needs of predictive models for overdose prevention, with the goal of supporting (1) safer prescribing of opioids (through a public health-level intervention), and (2) increasing the initiation of MOUDs (through an individual-level intervention).

Aim 1: Prospectively validate fatal and nonfatal opioid-related overdose prediction models using data from the Tennessee Department of Health.

Previously, Ripperger et al. used statewide data from Tennessee (2012-2017) to predict the risks of fatal and nonfatal opioid-related overdose within 30 days of a controlled substance prescription[9]. Models considered a combination of statewide prescription and hospitalization data in addition to ZIP-level socioeconomic indicators. The original study compared the performance of different ensembling methods and demonstrated that ridge regression ensembled models achieved the highest area under the receiver operating characteristic (fatal: 0.83; nonfatal: 0.83), area under the precision-recall curve (fatal: 0.00042; nonfatal: 0.0016), and calibration (Spiegelhalter p-value for fatal: 0.12; Spiegelhalter p-value for nonfatal: 0). However, since 2017, several substances have emerged in the Tennessee drug supply and contributed to overdose deaths[17–19]. Additionally, from 2018-2022, Black residents in Tennessee experienced a 325% increase in fatal overdoses, compared to an 80% increase among White residents[20]. A careful prospective validation of models from Ripperger et al. is necessary to understand how shifts in causal factors and cases might have impacted model performance[21].

Using Tennessee statewide data from 2022, we prospectively validated the fatal and nonfatal overdose prediction models developed by Ripperger[9]. We assessed model performance by region, sex, age group, rural-urban continuum codes, and prescription drug type. We identified high-risk prescriptions and aggregated prescription-level overdose predictions to the county level for comparison to observed overdose rates. We hypothesized that the previously developed models would show reduced performance on newer data, due to factors such as changes in the drug supply and overdose trends. This prospective validation is a pivotal

prerequisite for the operationalization of this model in the future. Additionally, this work could have important implications for the implementation and maintenance of other opioid-related overdose prediction models that use statewide data.

Aim 2: Develop and validate a predictive model for OUD medication response.

We developed and validated a model that predicted response to medication for OUD (MOUD), with the goal of informing treatment selection. Models used electronic health record (EHR) data from Vanderbilt University Medical Center (VUMC) and ZIP-level socioeconomic indicators from publicly available data sources. The cohort included patients who started medication for OUD in the hospital at VUMC from 2017-2022. We defined treatment success as the absence of (1) OUD medication utilization changes, (2) discharges against medical advice during the index hospitalization, and (3) substance use-related readmissions within 90 days of the index hospitalization. The model accounted for demographics (e.g., age, gender), social determinants (e.g., social deprivation index, insurance status, housing status), and clinical factors (e.g., history of substance use, psychiatric comorbidities). Findings from this study could inform OUD medication recommendations at the patient level.

Aim 3: Understand clinician perspectives on a predictive model for OUD medication.

While developing the model in Aim 2, we conducted semi-structured interviews with clinicians who prescribe MOUDs at VUMC. Interviews covered (1) current practices for introducing patients to MOUDs as treatment options, (2) current practices for selecting the specific MOUD, and (3) their perspectives on a hypothetical predictive model for OUD. Each interview was audio-recorded, transcribed using an online transcription service, and analyzed by

at least 2 investigators using software for qualitative data analysis. We conducted combined deductive-inductive content analysis based on a socio-technical model used for studying health information technology.[22] This work reveals socio-technical considerations that might impact the uptake of a predictive model in the context of OUD. Clinician perspectives might inform the development of other predictive models for OUD.

This work has advanced knowledge on the use of predictive models for overdose prevention in the domains of public health and clinical care. In the public health domain, we have validated a model that has the potential to inform the prevention of opioid-related overdose at the state level. In the clinical domain, we have developed a model for OUD treatment selection at the individual level. Additionally, we have gained a deeper understanding of how a predictive model for OUD medication might be implemented in the clinical setting.

1.2 Dissertation Roadmap

In this dissertation, Chapter 2 provides background information on opioids and the opioid epidemic. We present evidence-based approaches for overdose prevention and make the case for supporting a combination of interventions for reducing overdose deaths. We indicate that this dissertation will specifically focus on how predictive modeling might inform (1) safer opioid prescribing at the public health level and (2) treatment selection for OUD at the individual level.

At the public health level, we discuss the limitations of existing fatal and nonfatal overdose surveillance systems and the potential for predictive models to help facilitate more timely intervention. This sets the scene for Chapter 3, which describes work prospectively validating a predictive model that uses statewide data sources to predict opioid-related overdose.

At the individual level, we describe medications for OUD and factors that might make one medication preferable over the other in different situations. We highlight that there are no existing models that inform the decision between medications for OUD and shift our attention to major depressive disorder, for which treatment selection has been extensively studied and remains a challenge. Chapter 4 details our work adapting an existing modeling framework from major depressive disorder to OUD.

We also recognize that socio-technical factors are important to consider prior to implementation of any health information technology. As such, we interviewed clinicians to understand their perspectives on the treatment selection model described in Chapter 4 and how the model might be used at the point of care. This study is detailed in Chapter 5.

This dissertation ends with Chapter 6, in which we briefly recap the contributions from Chapters 3-5 and discuss the implications of this work for the domain of opioids as well as the field of biomedical informatics.

CHAPTER 2

BACKGROUND

2.1 A Primer on Opioids

Opioids are a class of pain-relieving medications. Opioids can be naturally derived from the opium poppy plant (e.g., morphine, codeine) or synthesized in laboratory settings (e.g., oxycodone, hydrocodone, fentanyl, buprenorphine)[23]. Prescription opioids such as hydrocodone, oxycodone, morphine, and codeine are commonly used to treat pain. While prescription opioids have been approved to treat pain, non-prescription opioids such as heroin and illicit fentanyl are far more potent and thus prohibited by law.

Both prescription and non-prescription opioids carry a high risk for addiction. This is because in addition to alleviating pain, opioids may cause euphoria. Individuals exposed to high doses of opioids or taking opioids for a prolonged period are at an increased risk for opioid dependence. Individuals with opioid dependence may experience withdrawal symptoms (e.g., pain, diarrhea and vomiting, cold flashes) while abstaining from opioids and may require higher doses of opioids to achieve the same narcotic effect. Opioid dependence becomes opioid use disorder (OUD) when opioid use significantly impacts daily life. OUD is a chronic mental health condition and is diagnosed using criteria outlined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)[24]. The DSM-5 defines OUD as repeated opioid use over the course of 12 months resulting in an individual meeting at least 2 of 11 criteria describing opioid craving, tolerance, withdrawal, and use. The severity of OUD is measured by the number of criteria observed within the 12-month period (mild: 2-3, moderate: 4-5, severe: 6 or more). Severe OUD is considered addiction.

Opioids can have severe physical effects, particularly if not used appropriately. Opioid misuse occurs if a prescription opioid is not used as directed or is shared with other individuals. Exposure to non-prescription opioids is also dangerous, owing to the increased potency of unregulated, lab-synthesized drugs such as illicit fentanyl. Prescription and non-prescription opioid use can result in adverse events such as drowsiness, confusion, nausea, constipation, and slowed breathing. The intake of life-threatening amounts of opioids can cause an overdose, which may result in brain damage, coma, or death.

2.2 The Ongoing Opioid Overdose Epidemic

A complex cascade of events contributed to the opioid epidemic, which was initially driven by prescription opioids. In 1995, the American Pain Society advocated for acute and cancer-related pain to be treated as a 5th vital sign[25]. Shortly after, in 2000, the Joint Commission and Veterans Health Administration published guidelines for pain management[26,27]. These standards encouraged physicians to treat pain, and patients to expect to be pain-free. When questions about pain management were added to surveys impacting reimbursements and quality of care ratings, physicians became incentivized to prescribe more opioids. Early studies suggested that opioids had low addiction potential in patient populations when prescribed appropriately[28]. At the same time, pharmaceutical companies misled the public on the addictive risks of manufactured opioids such as OxyContin; opioid distributors failed to report suspicious prescriptions[29–31].

The opioid overdose epidemic has occurred in three waves[1]. The first wave of the opioid overdose epidemic began in 1999 and involved primarily prescription opioids[32]. In 2010, a surge in heroin-related overdose deaths in 2010 marked the beginning of the second wave[33]. Shortly

after, in 2013, synthetic opioids such as fentanyl infiltrated the drug supply to mark the beginning of the third wave[34]. Heroin-related overdoses overtook prescription opioid-related overdoses in 2015[1]. Two years later in 2017, the opioid overdose epidemic was declared a public health emergency[35]. The rate of synthetic opioid-related deaths has continued to skyrocket, recording a 4% increase from 2021-2022[1]. In 2022, opioids contributed to 75% of overdose deaths[1].

2.3 Evidence-Based Interventions for Overdose Prevention

In 2022, opioid overdoses resulted in 81,806 deaths in the United States, the most recorded in any single year[1]. In response to the ongoing opioid overdose epidemic, the National Institutes of Health supported the Helping to End Addiction Long-term (HEAL) initiative. As part of this initiative, the HEALing Communities Study was launched in 2019 and aimed to identify interventions for reducing overdose deaths[36]. The HEALing Communities Study spanned 67 communities in 4 states heavily impacted by the epidemic: Kentucky, Massachusetts, New York, and Ohio. One outcome of this study was the Opioid-Overdose Reduction Continuum of Care Approach (ORCCA). ORCCA was developed by researchers in collaboration with community leaders, with the goal of informing the implementation of such interventions[2].

ORCCA organizes evidence-based practices for reducing overdose deaths into 3 categories. The first is preventing opioid-related overdoses through interventions such as naloxone distribution. Particularly as more overdoses involve multiple substances and synthetic drugs, additional doses of naloxone may be needed to reverse the effects of overdoses. This calls for increased awareness of where overdoses are occurring and where naloxone is distributed, to inform factors such as naloxone kit distribution and restocking. The second category is

facilitating the delivery of medication for OUD (MOUD). In 2022, only 25% of individuals with OUD reported receiving MOUD[10]. ORCCA suggests that more training and resources may be needed to ensure that prescribers and pharmacists are aware of prescribing regulations and comfortable initiating MOUDs. The third category is promoting safer opioid prescribing. Over 8 million people reported misusing prescription pain relievers, based on the 2023 National Survey on Drug Use and Health[37]. Factors such as the overprescribing of opioids for acute and postoperative pain, the prescribing of high doses (≥ 90 MME/day) or long durations (more than 7 days) of opioids, and the prescription of benzodiazepines in combination with opioids have contributed to overdose risk[38,39]. The enforcement of appropriate prescribing practices is important for reducing opioid misuse.

In a simulation study of the evidence-based practices described in ORCCA, Linas et al. found that no single intervention would reduce opioid-related overdoses by 40% in 2 years (from 2020 to 2022), a goal set by the United States Department of Health and Human Services[40]. Rather, a combination of 3 interventions would be necessary. Specifically, Linas et al. found that interventions would need to occur at the public health level (expanding naloxone distribution efforts) as well as the individual level (increasing MOUD initiation and retention respectively). Taken together, ORCCA and Linas et al. highlight the need for interventions that target individuals as well as those that impact the broader population. As such, this dissertation explores how we can inform the prevention of overdose deaths through both public health- and individual-level interventions to support 2 of the evidence-based practices in ORCCA: (1) safer prescribing of opioids (through a public health-level intervention) and (2) increasing the initiation of MOUDs (through an individual-level intervention).

2.4 Towards the Safer Prescribing of Opioids

2.4.1 High-Risk Opioid Prescribing

Prescription opioid misuse can occur if an individual does not take a prescription as directed (e.g., a higher dose or more often than needed) or takes medication prescribed to another individual. Over 8 million people reported misusing prescription opioids in 2023[37]. Most individuals reported receiving prescriptions either from a friend or family member (39.1%) or a health care provider (47%)[37]. Improving the safety of opioid prescribing involves reducing high-risk prescribing practices and enforcing appropriate prescribing practices, particularly for populations at increased risk for overdose.

ORCCA highlights three high-risk prescribing practices that should be reduced. First, higher doses of opioids have been linked to increased risks of unintentional fatal overdoses[41]. As a result, doses greater than 90 MME/day are considered high-risk. Second, the prescription of high-dose, long-acting opioids for opioid-naïve individuals is not recommended[39,42]. Third, the concurrent use of opioids and benzodiazepines has been linked to higher risks of respiratory depression and overdose[43].

Several clinical factors have been linked to increased overdose risk. Substance use history, concurrent use of opioids and other psychoactive drugs, psychiatric comorbidities (e.g., depression, bipolar, schizophrenia, anxiety), and certain clinical comorbidities (e.g., cirrhosis, renal insufficiency, congestive heart failure) are associated with increased risk for overdose[43]. Safer prescribing of opioids can be achieved by encouraging appropriate prescribing practices for populations at high risk of overdose.

2.4.2 Overdose Surveillance Systems

At the public health level, the United States Centers for Disease Control and Prevention (CDC) has established overdose surveillance systems for nonfatal and fatal drug overdose reporting. The CDC's Drug Overdose Surveillance Epidemiology (DOSE) System collects information on nonfatal overdoses[44]. DOSE relies on syndromic surveillance data, which includes 80% of emergency departments (ED) across the United States, and finalized hospital discharge data from ED and inpatient visits. Meanwhile, the State Unintentional Drug Overdose Reporting System (SUDORS) houses detailed information on unintentional fatal overdoses. SUDORS relies on data from a range of data sources including death certificates, toxicology reports, and coroner and medical examiner reports[45].

Both DOSE and SUDORS receive data from a total of 50 jurisdictions (49 states and the District of Columbia), each of which follows its own reporting regulations and laws. In fact, only 21 states had laws or regulations requiring nonfatal overdose reporting as of March 31, 2023[46]. Jurisdictions, in turn, receive data from local health departments, each of which follows its own reporting procedures. Procedures can rely on phone, fax, and manual review. The siloed data systems and time-consuming reporting methods contribute to reporting delays that can last months. This means that only overdose surveillance data from the previous year can be analyzed. Observational studies have contributed valuable information about how the opioid epidemic has impacted different communities, but delays in overdose reporting limit the utility of such studies for informing intervention.

2.4.3 Predictive Analytics Using Prescription Drug Monitoring Program Data

While overdose surveillance data are delayed, prescription drug monitoring programs (PDMPs) offer near real-time information on controlled substance prescriptions[47]. PDMPs were implemented to help reduce prescription opioid misuse and to facilitate controlled substance prescription oversight. PDMPs have been implemented in nearly every state and are accessible by pharmacists, clinicians, and state health departments. The CDC recommends that clinicians check the PDMP to understand controlled substance use history prior to prescribing opioids[38]. Predictive modeling on PDMP data could help identify high-risk prescribing and inform interventions promoting safer prescribing practices.

PDMP data have been used in overdose risk prediction models, but none of these predictive models have been prospectively validated to our knowledge[5–9,48]. Predictive models must be validated prior to implementation, especially in the case of overdose risk prediction. The opioid epidemic has disproportionately impacted certain racial and ethnic groups and regions across the nation[49–51]. From 2019-2020, racial and ethnic minority populations (non-Hispanic Black, non-Hispanic American Indian or Alaska Native) experienced the highest increases in overdose death rates based on data across 25 states[50]. In addition to ensuring that models are well calibrated and able to discriminate between cases and controls, it is important that model predictions and interventions are effective for all communities. It is critical that predictive models inform interventions that do not widen existing disparities, and prospective validation is one step towards safeguarding against this.

2.5 Towards Guiding Treatment Selection for Opioid Use Disorder

2.5.1 Treating Opioid Use Disorder with Medication

At the individual level, those with opioid use disorder (OUD) are at an increased risk for overdose. Over 9.3 million adults, or 3.7% of adults in the United States, were estimated to have OUD in 2022[10]. The first line of treatment for OUD is medication. Studies have shown that medications for OUD (MOUDs) are effective for reducing illicit opioid use and all-cause mortality[52–55]. Yet only 25% of individuals with OUD reported taking MOUD in 2022[10].

The United States Food and Drug Administration (FDA) has approved three MOUDs: naltrexone, methadone, and buprenorphine. Naltrexone is an antagonist, which means it acts by blocking other molecules from binding to opioid receptors altogether[56]. Naltrexone has been linked to lower patient acceptance, higher nonadherence, and worse outcomes and therefore has not been widely used for OUD[57–60]. As such, this dissertation will focus on methadone and buprenorphine as the standard treatment options for OUD.

Methadone was approved by the FDA for treatment of OUD in 1972[61]. Methadone is a full opioid receptor agonist that attaches to opioid receptors to produce a full opioid effect. At therapeutic doses, methadone provides pain relief, attenuates euphoria, and reduces withdrawal symptoms. Individuals do not need to be in withdrawal to begin methadone therapy. Contraindications to methadone include hypersensitivity, respiratory depression, severe asthma, and paralytic ileus.

Buprenorphine was approved for treatment of OUD in 2002. Buprenorphine is a partial agonist that does not completely bind to opioid receptors, resulting in a ceiling effect that reduces the risk of severe adverse events while still providing pain relief[62]. Individuals must be in mild withdrawal before starting buprenorphine therapy, to avoid precipitating withdrawal upon

initiation. Contraindications to buprenorphine include hypersensitivity, respiratory depression, and gastrointestinal obstruction.

2.5.2 Initiating Medication for OUD in the Hospital

In the hospital, clinicians are permitted to initiate individuals on either buprenorphine or methadone for OUD to achieve stabilization[61]. Studies have reported a wide range of MOUD retention rates, suggesting that treating OUD remains a challenge[11]. The decision between methadone and buprenorphine is complicated by factors such as insurance coverage, clinical comorbidities, access to methadone clinics, and patient preference[63–66].

To further complicate the MOUD decision, prescribing regulations are frequently changing and have impacted the accessibility of MOUDs. Buprenorphine became more accessible with the recent passing of the Mainstreaming Addiction Treatment (MAT) Act, which was signed on December 30, 2022. MAT enables any prescriber registered with the Drug Enforcement Administration to prescribe buprenorphine for OUD[67]. Prior to MAT, the Drug Abuse Treatment Act of 2000 prohibited clinicians without X-Waivers from prescribing buprenorphine for OUD[68]. Similarly, methadone has become more accessible with recent changes to prescribing regulations in April 2024[69]. These changes lower barriers to receiving larger supplies of methadone and allow additional clinicians like nurse practitioners and physician assistants to dispense methadone.

Studies have reported higher MOUD initiation rates when addiction psychiatry specialists have been involved in OUD care[12,13]. One study revealed that hospital-based clinicians believed they lacked training and experience in initiating MOUD[70]. Hospital-based clinicians who are not trained in addiction medicine may consult specialists like addiction psychiatrists to help lead the

decision between buprenorphine and methadone. However, as prescribing regulations continue to increase access to MOUDs, more clinicians will need to navigate the decision between MOUDs. One way to empower non-specialists to make quality care decisions without relying on specialists might be through a treatment selection model that provides personalized MOUD recommendations.

2.5.3 Treatment Selection Models for More Personalized Care in Mental Health

Buprenorphine and methadone have been well studied in clinical trials and observational studies. Retention rates slightly favor methadone, but the two MOUDs appear to be relatively equivalent in secondary outcomes such as reduced cravings and use of other substances[71]. Previous studies have typically involved only one MOUD or have identified risk factors for treatment discontinuation. Factors such as age, gender, substance use, employment status, race, MOUD dose, and insurance status have been linked to early treatment discontinuation[72–77]. Predictive modeling has been used to predict early dropout from OUD treatment[14–16]. However, no studies to our knowledge have attempted to inform the decision between MOUDs.

Treatment selection has been extensively studied for major depressive disorder (MDD), a mental health disorder which is like OUD in many ways. MDD is a mood disorder characterized by symptoms such as sadness, angry outbursts, loss of interest in hobbies, and feelings of worthlessness[78]. MDD and OUD share similar neurobiology and co-occur[79]. MDD can be chronic or manifest in acute episodes of relapse that interrupt periods of remission. As with OUD treatment, stopping MDD treatment can result in physical dependence or withdrawal-like symptoms. Treatment retention is a challenge for both OUD and MDD[11,80].

Many treatment options exist for MDD. There are several FDA-approved medications for MDD, such as selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, mirtazapine, and bupropion. Psychotherapy-based approaches such as cognitive behavioral therapy and somatic treatments such as ECT have also been proven to be effective[81,82]. A significant amount of trial and error is often needed before the appropriate treatment for an individual is identified[83].

A treatment selection model that leverages patient characteristics to predict treatment response could help clinicians make more personalized care decisions for MDD. Various combinations of genetic, sociodemographic, and clinical data have been used in machine learning models for predicting MDD treatment response[84]. However, most models have relied on limited data describing biomarkers, MDD symptom scales, and studies designed specifically for MDD.

Sheu et al. have been the first to our knowledge to use real-world electronic health record (EHR) data to predict antidepressant treatment response[85]. Models from Sheu et al. showed good performance and can be used to generate probabilities of response for each antidepressant class. Additionally, Sheu et al. present an interpretable visualization that highlights top predictors for each treatment option. The use of widely available data makes the approach taken by Sheu et al. reproducible at other institutions and potentially generalizable to other mental health conditions. In the case of OUD, a treatment selection model like the one presented by Sheu et al. could empower non-specialists to make treatment decisions like specialists. However, the modeling framework has not yet been tested on MOUDs.

2.5.4 Challenges of Implementing Predictive Models as Clinical Decision Support

Clinical decision support (CDS) systems include computerized alerts, order sets, and dashboards. One way that CDS systems have been categorized is knowledge-based vs. nonknowledge-based (i.e., data-driven) CDS systems[86]. Knowledge-based CDS systems function to help clinicians arrive at care decisions and typically use rule-based programs. On the other hand, data-driven CDS systems offer recommendations derived from system-generated rules and can leverage artificial intelligence (AI)-based algorithms. This dissertation focuses on the latter, which is particularly important given the rising popularity of AI in medicine and beyond.

CDS tools have been developed for a variety of tasks that include providing diagnostic support, medication order support, and drug interaction alerts. One challenge of CDS systems is clinician uptake. Across 55 CDS studies, overall clinician uptake of CDS tools was only 34.2%[87]. Clinician uptake of CDS has been linked to user- and system-specific factors[88,89]. Important factors at the user level include user perception of CDS tools in general and a retained sense of autonomy. Important factors at the system level include integration with the clinical workflow and the EHR, in addition to the usability of the tool. For data-driven CDS specifically, additional challenges for uptake include the lack of interpretability of black box models and understanding how models should be maintained and updated as data change over time[90].

Achieving good enough predictive performance is only one prerequisite for implementing a predictive model as CDS. Prior to implementation, it is important to consider the "five rights" of CDS in the context of the specific CDS system: the right information, the right people, the right format, the right channel, and the right time in a clinical workflow[91]. Sittig and Singh expand on this framework in a socio-technical model designed to help address challenges related to health information technology[22]. The socio-technical model details 8 dimensions that describe

the interaction of people, processes, and technology within internal and external contexts. The socio-technical needs of clinicians using AI-based predictive models were explored by Benda et al. in a pre-implementation study of a predictive model for identifying patients with high preventable utilization patterns[92]. In their study, Benda et al. found that socio-technical considerations were more important than purely technical considerations. The case of OUD treatment selection is particularly compelling and worth studying specifically, given the impact of social factors such as access to care as well as external factors such as prescribing regulations.

2.6 Dissertation Overview

This dissertation applies predictive modeling to inform overdose prevention at both the public health and individual levels. To promote safer opioid prescribing practices at the public health level, we prospectively validated overdose risk prediction models from Ripperger et al[9]. The models used data from the state of Tennessee, which recorded the third highest rate of overdose deaths in 2022[3]. To increase initiation of MOUDs at the individual level, we developed and validated OUD treatment recommendation models adapted from Sheu et al[85]. At the same time, we interviewed clinicians to understand the socio-technical factors impacting their perceptions of a predictive model recommending medication for OUD. Together, this work supports the development of public health- and individual-level predictive modeling-based interventions for overdose prevention.

CHAPTER 3

PROSPECTIVE VALIDATION OF ENSEMBLE LEARNING MODELS FOR PREDICTING OPIOID-RELATED OVERDOSE USING TENNESSEE STATEWIDE DATA

3.1 Introduction

In 2022, a total of 81,806 opioid-related overdose deaths occurred, the highest number recorded in any single year in the United States[1]. While the opioid epidemic was initially driven by prescription opioids, heroin-related overdoses overtook prescription opioid-related overdoses in 2015[1,32]. The rate of synthetic opioid-related deaths has continued to skyrocket, recording a 4% increase from 2021-2022[1]. In 2022, opioids were linked to 75% of overdose deaths[1].

One challenge of overdose prevention is obtaining timely surveillance data to inform interventions. Each state follows its own death reporting process, which can be delayed by factors related to the type of death investigation system (e.g., centralized medical examiner office, county-based coroner) or the extent of electronic death registration system adoption[4]. While observational studies on statewide data have elucidated trends in and risk factors for overdose deaths, predictive models can inform more timely intervention[1,93,94]. The prescription drug monitoring program (PDMP) was established to track controlled substance prescriptions and has been implemented in nearly all jurisdictions in the United States[95]. The PDMP offers comprehensive, near real-time data for overdose risk prediction.

Models for overdose prediction have used PDMP data. An existing proprietary algorithm generates overdose risk scores and supports 80% of statewide PDMP systems[48]. However, the details of this algorithm have not been disclosed, and there is no evidence that this model has

been clinically validated with relevant care outcomes[96,97]. Other predictive models for overdose have been developed at the state level using PDMP and statewide data from Maryland, Oregon, and Pennsylvania[5–8]. Ripperger et al. were the first to use statewide data for opioid-related overdose prediction in a southern state[9]. None of these models have been prospectively validated to our knowledge. Little is known about the generalizability of an overdose prediction model developed using statewide data from nearly ten years prior.

In addition to ensuring that predictive models are well calibrated and able to discriminate between cases and controls, it is important that model predictions and corresponding interventions are accurate and effective across all communities. The opioid epidemic has disproportionately impacted racial and ethnic minority populations and regions across the nation[49–51]. Shifts in the prevalences of cases and underlying causal factors present additional challenges to the transportability of models across data spanning different states and time periods[21]. It is critical that predictive models do not widen existing disparities, and prospective validation is one step towards safeguarding against this.

The fatal and nonfatal overdose prediction models developed by Ripperger et al. using data from Tennessee are particularly compelling, because Tennessee recorded the third highest overdose death rate across the United States in 2022[3]. In Tennessee, emerging substances such as xylazine, nitazine, and tianeptine have continued to contaminate the local drug supply and contribute to overdose deaths[17–19]. From 2018-2022, Black residents experienced a 325% increase in fatal overdoses, compared to an 80% increase among White residents[20]. The dynamic nature of the opioid epidemic and its disproportionate impact on racial and ethnic minority populations calls for careful validation of predictive models that may inform public health

interventions. In this study, we evaluated the performance of fatal and nonfatal opioid-related overdose prediction models from Ripperger et al. using Tennessee statewide data from 2022.

3.2 Methods

Previously in Ripperger et al. 2022 (hereafter referred to as the original study), ensemble learning models predicting fatal and nonfatal opioid-related overdose risk were developed and validated using Tennessee statewide data (2012-2017)[9]. For a given controlled substance prescription, the models each output a predicted likelihood of a (fatal | nonfatal) opioid-related overdose occurring within 30 days of the prescription. Model predictors included prescription- and ZIP-level factors derived from statewide and publicly available data sources described below. The original study compared the performance of ensembled models and demonstrated that the ridge regression ensembled models achieved the highest discrimination and calibration. The fatal overdose risk prediction model recorded an area under the precision-recall curve (AUPRC) of 0.0042, an area under the receiver operating characteristic (AUROC) of 0.83, and was calibrated (Spiegelhalter p-value = 0.12). The nonfatal overdose risk prediction model recorded an AUPRC of 0.0016, an AUROC of 0.83, and was miscalibrated (Spiegelhalter p-value = 0).

In this prospective validation study, the ridge regression ensembled models that were trained in the original study were evaluated using Tennessee statewide data from 2022. This work was only possible through an academic-state partnership between the Tennessee Department of Health (TDH) and Vanderbilt University Medical Center (VUMC). The Vanderbilt University Institutional Review Board (#171323) approved this study.

3.2.1 Data Sources

This study used the same data sources as the original study. At the TDH, the Integrated Data System (IDS) contains statewide data linked across several databases[100]. The prescription drug monitoring program (PDMP) records information on controlled substance prescriptions. Vital records include data from birth, death, marriage, and divorce certificates. The Hospital Discharge Data System (HDDS) contains information on inpatient hospitalizations, emergency department visits, and select outpatient records. The State Unintentional Drug Overdose Reporting System (SUDORS), which contains manually curated details on unintentional fatal overdoses, also contributes data to the IDS. Tennessee statewide data spanned the study period of April 4, 2022 - December 1, 2022. The dataset was limited to prescriptions for individuals who were between the ages of 20 and 89 and resided within the state of Tennessee at the time of prescription fill.

This study also included data from publicly available databases. As described by Ripperger et al., statewide data were combined with socioeconomic indicators available from public data sources. Socioeconomic indicators were at the ZIP level or mapped from counties to ZIP codes encapsulating the majority area of the county. Data sources included the American Community Survey (ACS), TDH, United States Economic Census (USEC), United States Centers for Disease Control (CDC), United States Department of Agriculture (USDA), and University of Wisconsin (UW). Socioeconomic indicators were not updated from the original study. The most recent values from the original study were used. For example, if ACS results were available for each year from 2012 to 2017, the value from 2017 was included.

3.2.2 Outcomes

Fatal and nonfatal opioid-related overdoses occurring within 30 days of prescriptions were identified following TDH surveillance methods[101,102]. Fatal opioid-related overdoses were identified using data from SUDORS, and nonfatal opioid-related overdoses (excluding heroin) were identified using data from the HDDS. Nonfatal outcomes were derived using the same methods as the original study, while fatal outcomes were limited to *unintentional* fatal overdoses, which comprised 99.5% of all fatal overdoses in 2022[20,102].

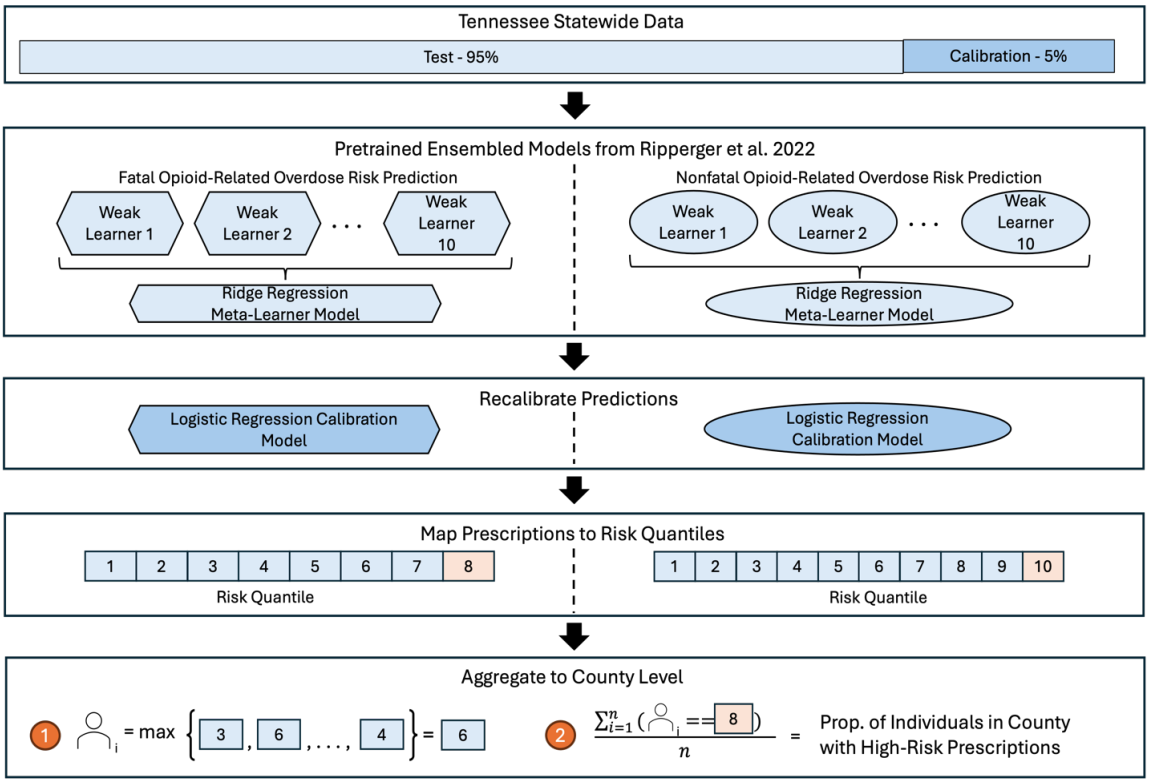
3.2.3 Model Predictors

As in the original study, the model had a total of 342 predictors (**Appendix A – Figure 10**). For each prescription recorded in the PDMP during the study period, predictors occurring within 1 year prior were considered. Variables from the PDMP and HDDS were grouped using National Drug File-Reference terminology (NDF-RT), Pharmacologic Classes and Clinical Classification Software (CCS), Level 2 groupings from National Drug Codes (NDCs) and International Classification of Disease, revision 10, Clinical Modification (ICD-10-CM) codes. Predictors from the PDMP and HDDS included sex, age, number of unique practitioners, and unique hospital identifiers. Total prior PDMP prescriptions, morphine milligram equivalents prescribed, prescriptions of various medications (e.g., benzodiazepines, central nervous system stimulant, opioid agonist, overlapping opioid and benzodiazepine prescriptions), hospital visits, and hospital discharge diagnoses of different types (e.g., spondylosis, substance-related disorders, hypertension, mood disorders) were also considered.

ZIP-level predictors were derived from publicly available data sources. Variables from the TDH included the number of pain management clinics within a 60-mile radius, the age

adjusted rate of inpatient stays involving drug overdose, the age adjusted rate of outpatient stays involving heroin overdose, and the Tennessee human immunodeficiency virus (HIV) or hepatitis C virus (HCV) transmission vulnerability index. Rural-urban continuum (RUCC) codes (minimum, maximum) from the USDA, area deprivation index (minimum, maximum, average, median) from UW, number of employees (USEC), median household income (USEC), and social vulnerability indices (CDC) were included. Information on factors such as the percentage of households under poverty line, single parent households, employed population, individuals with some level of college education, and privately insured were sourced from the ACS.

Figure 1: Overdose risk prediction model recalibration and prediction aggregation steps.



3.2.4 Model Recalibration

The fatal and nonfatal overdose prediction models followed the same recalibration steps (**Figure 1**). Statewide data were randomly split into calibration (5%) and test sets (95%). Models were recalibrated using Platt scaling, which involved training a logistic regression model on the predicted probabilities from the calibration set and applying the model to predictions from the test set.

3.2.5 Assessment of Model Performance at the Prescription Level

The same model performance metrics were calculated for the fatal and nonfatal overdose prediction models. Across prescription-level predictions, model performance metrics were calculated using 95% bootstrapped confidence intervals. The AUROC and AUPRC were used to evaluate model discrimination. The Brier score and Spiegelhalter z-test were used to evaluate model calibration. A risk concentration analysis was performed using 8 quantiles for the fatal and 10 quantiles for the nonfatal models. These were the same quantiles for which results were reported in the original study[9]. The risk concentration analysis involved calculating the proportion of cases captured by each risk quantile and the number needed to screen (NNS) to identify one case in each risk quantile. Fatal fentanyl-related overdoses captured by each quantile were also noted. Fentanyl-related overdoses were identified consistent with TDH surveillance methods[103].

Model performance was evaluated at the prescription level across different subgroups. Subgroups included grand region of Tennessee (east, middle, west), age group (20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89), sex (male, female, unknown), rural-urban continuum codes (RUCC: metro, urban, rural), and PDMP prescription drug class (benzodiazepine, gabapentin,

miscellaneous zolpidem, muscle relaxant, opioid, stimulant, testosterone, unknown). The proportion of cases and controls, AUROC, and AUPRC were calculated for each subgroup.

3.2.6 Aggregation of Prescription-Level Predictions to the County Level

Predictions from the fatal and nonfatal models were aggregated to the county level in parallel, following the same steps outlined in **Figure 1**. Each prescription-level prediction was mapped to the county level using the county of residence recorded in the PDMP at the time of prescription fill. For each county, an aggregated county-level prediction was derived by calculating the proportion of prescriptions that were high risk. High-risk prescriptions were defined as prescriptions with predictions corresponding to the highest quantiles from the risk concentration analysis (8 for the fatal overdose model; 10 for the nonfatal overdose model). For individuals who filled multiple prescriptions, only the highest prediction was considered for each county of residence. Therefore, the same individual could contribute to multiple county-level predictions but would only be represented once in each county.

3.2.7 Comparison of Aggregated County-Level Predictions to Observed Overdose Rates

Aggregated county-level predictions were ranked and compared to observed rates of fatal and nonfatal opioid-related overdoses in 2022. Observed overdose rates were extracted from the Tennessee Drug Overdose Dashboard[104]. Overdose rates were only available for counties reporting more than 20 cases. Nonfatal opioid-related overdose rates included inpatient and outpatient visits. County-level predictions and observed overdose rates for each model were overlayed on a map of Tennessee. For fatal overdoses, discrepancies between the counties of

residence at time of death based on SUDORS and the counties of residence at prescription fill based on the PDMP were noted and reported as a percentage.

3.3 Results

3.3.1 Cohort

A total of 6,386,989 controlled substance prescriptions were filled from April 4, 2022 to December 1, 2022. Approximately 2% of prescriptions (n=129,276 across 32,565 individuals) were removed because they were filled by individuals less than 20 years of age or greater than 89 years of age. Forty-two prescriptions across 38 individuals were excluded because they were linked to counties that were either in a different state (n=33 prescriptions across 32 individuals), unknown in Tennessee (n=8 prescriptions across 5 individuals), or missing (n=1 prescription).

The final dataset contained 6,257,671 controlled substance prescriptions across 1,319,573 individuals. At the individual level, 53% were female, 34% were male, and 13% were unknown. Individuals had a median of 3 prescriptions (interquartile range: [1, 7]) over the study period. At the prescription level, 39% of prescriptions were opioids, 19% were benzodiazepines, 13% were gabapentin, 13% were stimulants, and 16% were other drugs. East Tennessee had the most prescriptions (42%), followed by Middle Tennessee (38%), West (20%), and unknown (0.22%).

3.3.2 Outcomes

Opioid-related fatal overdoses occurred within 30 days of 506 prescriptions (0.0081%) across 314 individuals (0.024%), and nonfatal overdoses occurred within 30 days of 2,787 prescriptions (0.045%) across 1,606 individuals (0.122%) (**Table 1)**. In the calibration set of

312,883 prescriptions (5%) across 261,743 individuals (20%), opioid-related fatal overdoses occurred within 30 days of 22 prescriptions (0.007%) across 21 individuals (0.008%). Nonfatal overdoses occurred within 30 days of 154 prescriptions (0.049%) across 149 individuals (0.057%). In the test set of 5,944,788 prescriptions (95%) across 1,299,812 individuals (99%), opioid-related fatal overdoses occurred within 30 days of 484 prescriptions (0.0081%) across 304 individuals (0.023%). Nonfatal overdoses occurred within 30 days of 2,633 prescriptions (0.044%) across 1,551 individuals (0.12%).

Table 1: Prospective validation results for ridge regression ensembled models.

The dataset in Ripperger et al. 2022 contained 71,479,191 prescriptions across 3,041,668 individuals. The dataset in this prospective validation contained 6,257,671 prescriptions across 1,319,573 individuals.

	Fatal		Nonfatal	
	Ripperger et al. 2022	**Prospective Validation**	**Ripperger et al. 2022**	**Prospective Validation**
Outcomes[a]				
Prescriptions	4,912 (0.0069%)	506 (0.0081%)	19,460 (0.027%)	2,787 (0.045%)
Individuals	2,574 (0.085%)	314 (0.024%)	8,455 (0.28%)	1,606 (0.12%)
Discrimination				
AUROC	0.83	0.83 [0.81, 0.84]	0.83	0.76 [0.75, 0.77]
AUPRC	0.00042	0.00157 [0.0015, 0.003]	0.0016	0.0016 [0.0015, 0.0018]
Calibration				
Brier Score	6.73e-05	8.2e-05 [7.5e-05, 8.9e-05]	2.9e-04	4.4e-04 [4.3e-04, 4.7e-04]
Intercept	-0.33	1.9 [0.33, 3.3]	-1.75	-0.76 [-0.98, -0.47]
Slope	0.95	1.2 [1, 1.3]	0.79	0.91 [0.88, 0.95]
Spiegelhalter z-score	1.55	3.4 [1.4, 5.4]	-9.34	-5.3 [-7.1, -2.9]
Spiegelhalter p-value	0.12	7e-04 [5.9e-08, 0.17]	0	1.1e-7 [1.4e-12, 0.0042]

[a] Across the entire dataset

AUROC: area under the receiver operating characteristic; AUPRC: area under the precision-recall curve.

3.3.3 Model Performance Assessment

Compared to the original study, the fatal opioid-related overdose prediction model recorded a similar AUROC, a higher AUPRC, and a higher Brier score. The model became miscalibrated in the prospective validation and remained miscalibrated after Platt scaling.

Compared to the original study, the nonfatal opioid-related overdose prediction model recorded a lower AUROC, similar AUPRC, and higher Brier score in the validation study. The model remained miscalibrated after Platt scaling.

3.3.4 Risk Concentration Analysis

For the fatal overdose prediction model, the top risk quantile captured 54% (n=260) of the 484 prescriptions linked to opioid-related fatal overdoses and 53% (n=191) of the 363 prescriptions linked to fentanyl-related fatal overdoses in the test set **(Table 2)**. The NNS to identify 1 opioid-related fatal overdose was 2,859 for the top risk quantile bin. Each of the 8 quantiles captured between 743,097 and 743,100 prescriptions.

For the nonfatal overdose prediction model, the top risk quantile captured 37% (n=964) of the 2,663 prescriptions linked to nonfatal overdoses, and the top 2 risk quantiles together contained over half of the prescriptions linked to nonfatal overdoses. The NNS for the top risk quantile bin was 617. Each of the 10 quantiles contained between 594,477 and 594,481 prescriptions.

3.3.5 Subgroup Analysis

Both models recorded higher AUROCs for females than males (**Table 3**). The nonfatal overdose risk prediction model recorded significantly lower AUROCs for older age groups. Both

models showed decreased performance for prescription opioids when compared to benzodiazepine and gabapentin.

Table 2: Risk concentration of ensembled fatal and nonfatal model predictions in test set.

The test set contained 5,944,788 prescriptions across 1,299,812 individuals.

	Quantile	Prescriptions	Opioid-related Overdoses[a]	Fentanyl-related Overdoses[b]	NNS[c]	Quantile Range
Fatal	1	743,097	5 (0.01)	5 (0.014)	148,620	[7.0113e-05,7.0119e-05)
	2	743,100	5 (0.01)	5 (0.014)	148,620	[7.0119e-05,7.0121e-05)
	3	743,098	7 (0.014)	6 (0.017)	106,157	[7.0121e-05,7.0123e-05)
	4	743,099	8 (0.017)	3 (0.0083)	92,888	[7.0123e-05,7.0127e-05)
	5	743,097	42 (0.087)	29 (0.08)	17,693	[7.0127e-05,7.0133e-05)
	6	743,100	67 (0.14)	52 (0.14)	11,092	[7.0133e-05,7.0147e-05)
	7	743,097	90 (0.19)	72 (0.2)	8,257	[7.0147e-05,7.0188e-05)
	8	743,100	260 (0.54)	191 (0.53)	2,859	[7.0188e-05,0.014601)
Nonfatal	1	594,479	28 (0.011)	-	21,232	[4.453e-04,4.455e-04)
	2	594,478	52 (0.02)	-	11,433	[4.455e-04,4.456e-04)
	3	594,480	79 (0.03)	-	7,526	[4.456e-04,4.457e-04)
	4	594,478	84 (0.032)	-	7,078	[4.457e-04,4.459e-04)
	5	594,479	124 (0.047)	-	4,795	[4.459e-04,4.464e-04)
	6	594,479	180 (0.068)	-	3,303	[4.464e-04,4.472e-04)
	7	594,478	265 (0.1)	-	2,244	[4.472e-04,4.492e-04)
	8	594,477	342 (0.13)	-	1,739	[4.492e-04,4.551e-04)
	9	594,481	515 (0.2)	-	1,155	[4.551e-04,4.805e-04)
	10	594,479	964 (0.37)	-	617	[4.805e-04,5.031e-02)

[a] Number of opioid-related overdoses (proportion of opioid-related overdoses); includes fentanyl-related overdoses
[b] Number of fentanyl-related overdoses (proportion of fentanyl-related overdoses)
[c] Number needed to screen to identify 1 opioid-related overdose

Table 3: Model metrics for overdose risk prediction models, by subgroup.

Median and 95% bootstrapped confidence intervals were calculated for the test set (n = 5,944,788 prescriptions across 1,299,812 individuals).

Subgroup		Fatal				Nonfatal			
		AUPRC	AUROC	Cases n=484	Controls n=5,944,304	AUPRC	AUROC	Cases n=2,633	Controls n=5,942,155
Grand Region	**East**	0.0029 [0.0026, 0.0057]	0.81 [0.79, 0.84]	219 (45.25%)	2489498 (41.97%)	0.0017 [0.0016, 0.0022]	0.75 [0.74, 0.76]	998 (37.9%)	2488719 (41.97%)
	Middle	0.004 [0.0034, 0.0076]	0.85 [0.82, 0.88]	170 (35.12%)	2235034 (37.68%)	0.002 [0.0019, 0.0025]	0.76 [0.75, 0.77]	1105 (41.97%)	2234099 (37.68%)

	West	0.0063 [0.0048, 0.013]	0.82 [0.79, 0.85]	95 (19.63%)	1206973 (20.35%)	0.0026 [0.0023, 0.0037]	0.78 [0.77, 0.8]	530 (20.13%)	1206538 (20.35%)
Sex	**Female**	0.0036 [0.0027, 0.0077]	0.83 [0.81, 0.85]	222 (45.87%)	3260931 (54.98%)	0.0018 [0.0017, 0.0022]	0.75 [0.74, 0.76]	1570 (59.63%)	3259583 (54.97%)
	Male	0.0028 [0.0025, 0.0049]	0.79 [0.77, 0.81]	257 (53.1%)	2028879 (34.21%)	0.0018 [0.0016, 0.0023]	0.72 [0.7, 0.73]	1047 (39.76%)	2028089 (34.2%)
	Unknown	0.13 [0.056, 0.5]	0.93 [0.87, 0.96]	5 (1.03%)	641695 (10.82%)	0.039 [0.021, 0.084]	0.92 [0.85, 0.98]	16 (0.61%)	641684 (10.82%)
Age Group	**20-29**	0.015 [0.011, 0.031]	0.87 [0.83, 0.9]	41 (8.47%)	302031 (5.09%)	0.0067 [0.0058, 0.011]	0.83 [0.79, 0.86]	117 (4.44%)	301955 (5.09%)
	30-39	0.0065 [0.0047, 0.014]	0.84 [0.81, 0.87]	105 (21.69%)	700642 (11.81%)	0.0038 [0.0034, 0.0062]	0.83 [0.8, 0.85]	258 (9.8%)	700489 (11.81%)
	40-49	0.014 [0.0045, 0.036]	0.8 [0.77, 0.84]	125 (25.83%)	872909 (14.72%)	0.003 [0.0027, 0.0045]	0.81 [0.79, 0.83]	379 (14.39%)	872655 (14.72%)
	50-59	0.006 [0.0052, 0.012]	0.87 [0.84, 0.9]	101 (20.87%)	1115325 (18.8%)	0.0026 [0.0023, 0.0038]	0.78 [0.76, 0.8]	461 (17.51%)	1114965 (18.8%)
	60-69	0.0081 [0.0066, 0.017]	0.82 [0.79, 0.86]	67 (13.84%)	1205180 (20.32%)	0.0023 [0.0021, 0.0033]	0.74 [0.73, 0.76]	614 (23.32%)	1204633 (20.32%)
	70-79	0.063 [0.029, 0.17]	0.57 [0.36, 0.83]	10 (2.07%)	824232 (13.9%)	0.0027 [0.0024, 0.004]	0.7 [0.68, 0.73]	400 (15.19%)	823842 (13.89%)
	80-89	-	-	-	340103 (5.73%)	0.0041 [0.0037, 0.0074]	0.63 [0.59, 0.67]	155 (5.89%)	339948 (5.73%)
Rural-Urban Continuum Code Group	**Metro (1-3)**	0.002 [0.0018, 0.0035]	0.83 [0.81, 0.85]	397 (82.02%)	4347419 (73.29%)	0.0017 [0.0016, 0.002]	0.77 [0.76, 0.78]	1962 (74.52%)	4345854 (73.29%)
	Urban (4-7)	0.0075 [0.0058, 0.015]	0.81 [0.78, 0.85]	80 (16.53%)	1406238 (23.71%)	0.0021 [0.0019, 0.003]	0.73 [0.7, 0.75]	591 (22.45%)	1405727 (23.71%)

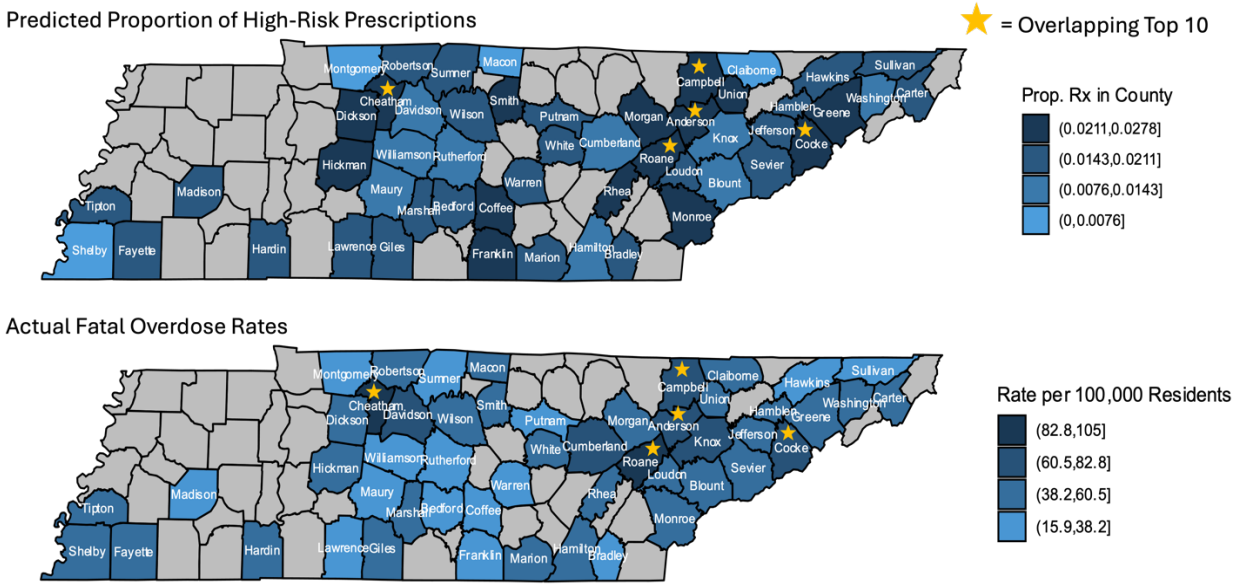
	Rural (8-9)	0.1 [0.042, 0.25]	0.65 [0.51, 0.83]	7 (1.45%)	177848 (3%)	0.0095 [0.0078, 0.018]	0.81 [0.77, 0.85]	80 (3.04%)	177775 (3%)
Prescription Drug Class	**Benzodiazepine**	0.0062 [0.0051, 0.013]	0.83 [0.8, 0.86]	96 (19.83%)	1138679 (19.2%)	0.0027 [0.0025, 0.0041]	0.8 [0.78, 0.82]	380 (14.43%)	1138395 (19.2%)
	Gabapentin	0.0082 [0.0064, 0.016]	0.84 [0.8, 0.87]	70 (14.46%)	797297 (13.44%)	0.003 [0.0028, 0.0046]	0.79 [0.77, 0.81]	376 (14.28%)	796991 (13.44%)
	Opioid	0.0031 [0.0025, 0.006]	0.79 [0.76, 0.81]	251 (51.86%)	2287552 (38.57%)	0.0018 [0.0017, 0.0022]	0.68 [0.66, 0.69]	1598 (60.69%)	2286205 (38.56%)
	Stimulant	0.023 [0.015, 0.048]	0.9 [0.87, 0.94]	26 (5.37%)	762387 (12.85%)	0.02 [0.0068, 0.074]	0.86 [0.83, 0.9]	77 (2.92%)	762336 (12.86%)
	Testosterone	0.13 [0.056, 0.33]	0.93 [0.87, 0.98]	5 (1.03%)	149852 (2.53%)	0.1 [0.046, 0.25]	0.79 [0.6, 0.95]	6 (0.23%)	149851 (2.53%)
	Miscellaneous Zolpidem	0.043 [0.025, 0.096]	0.81 [0.74, 0.9]	15 (3.1%)	301461 (5.08%)	0.011 [0.0089, 0.022]	0.85 [0.81, 0.9]	57 (2.16%)	301419 (5.08%)
	Muscle Relaxant	-	-	-	13686 (0.23%)	0.086 [0.045, 0.25]	0.72 [0.41, 0.93]	7 (0.27%)	13679 (0.23%)
	Other	0.027 [0.018, 0.059]	0.83 [0.75, 0.91]	21 (4.34%)	480591 (8.1%)	0.0051 [0.0045, 0.0097]	0.8 [0.76, 0.83]	132 (5.01%)	480480 (8.1%)

3.3.6 Comparison of County-Level Predictions to Observed Overdose Rates

Both models approximated the general trend of actual overdose rates. For fatal overdoses, 5 of the top 10 counties with the highest observed overdose rates were also in the top 10 counties with highest county-level predictions (**Figure 2**). County of residence at the time of prescription aligned with the county of residence at time of death for 85% (n=409) of cases. For nonfatal overdoses, 6 of the top 10 counties with the highest observed overdose rates were also in the top 10 counties with highest county-level predictions (**Figure 3**).

Figure 2: Fatal Overdose Model – Comparison of predicted high-risk prescriptions vs. observed fatal overdose rates in 2022.

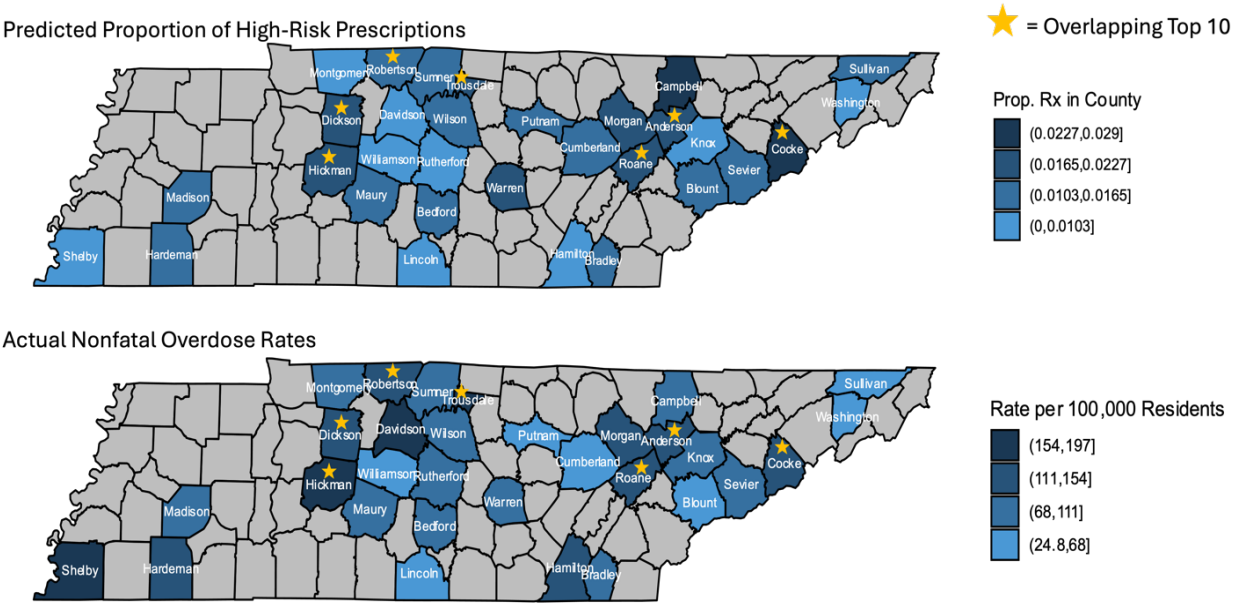
When comparing predicted proportions of high-risk prescriptions to observed fatal overdose rates, 5 of the top 10 counties overlapped. Only counties with more than 20 cases are shown.



County	Predicted Rank	Observed Rank	Overlapping Top 10
Campbell	1	3	x
Cocke	2	8	x
Smith	3	11.5	
Anderson	4	6	x
Dickson	5	22	
Franklin	6	43	
Cheatham	7	2	x
Coffee	8	40	
Morgan	9	18	
Roane	10	1	x

Figure 3: Nonfatal Overdose Model - Comparison of predicted high-risk prescriptions vs. observed nonfatal overdose rates in 2022.

When comparing predicted proportions of high-risk prescriptions to observed nonfatal overdose rates, 7 of the top 10 counties overlapped. Only counties with more than 20 cases are shown.



County	Predicted Rank	Observed Rank	Overlapping Top 10
Cocke	1	6	x
Campbell	2	13	
Trousdale	3	3	x
Morgan	4	12	
Dickson	5	5	x
Hickman	6	2	x
Anderson	7	8	x
Roane	8	9.5	x
Warren	9	20	
Robertson	10	7	x

3.4 Discussion

Ripperger et al. were the first to use statewide data from a southern state for models predicting opioid-related overdose risk within 30 days of a controlled substance prescription.

Models were originally developed and validated using Tennessee statewide data from 2012-2017. Using a combination of statewide data (PDMP, HDDS) and county-level socioeconomic indicators, models predicted fatal and nonfatal opioid-related overdose risk at the prescription level. In this study, fatal and nonfatal opioid-related overdose prediction models from Ripperger et al. were prospectively validated on Tennessee statewide data from 2022. Prescription-level predictions were analyzed by subgroup and aggregated to the county level to facilitate public health decision-making. Aggregated county-level predictions were compared to observed overdose rates.

3.4.1 Models Performed Similarly on Newer Data

Compared to the original study, the validation dataset had higher prevalences of prescriptions linked to fatal and nonfatal overdoses respectively. Additionally, overdose deaths have increasingly involved stimulants, benzodiazepines, and other emerging substances[20]. Despite these differences, the models did not need retraining or updating to record discrimination metrics comparable to those reported in the original study. The fatal opioid-related overdose prediction model achieved a similar AUROC and a higher AUPRC. The nonfatal opioid-related overdose prediction model recorded a slightly lower AUROC and a similar AUPRC. Notably, similar model performance was achieved even though publicly available socioeconomic indicators were not updated from the original study.

3.4.2 Models Needed Recalibrating

An off-the-shelf recalibration method (Platt scaling) did not achieve calibration. Shifts in the prevalences of cases and causal factors (e.g., prescribing regulations, drug supply) over time

likely contributed to the observed miscalibration[105]. Perhaps a larger, more representative calibration dataset or a more complex recalibration method was needed. Future work might involve measuring the shifts in data between the original and prospective validation studies, as well as characterizing the groups for which the model over- and under-predicted risk. We note that the proposed method for identifying high-risk prescriptions may not require a perfectly calibrated model, since the prescriptions corresponding to the highest risk quantile are unlikely to change after calibration. However, use cases that involve setting a threshold for identifying overdose cases will require a well-calibrated model.

3.4.3 Identifying High-Risk Prescriptions from Model Predictions

Model predictions were developed at the prescription level and can be aggregated to different levels to inform intervention. We prospectively validated a risk concentration approach that can be used to identify high-risk prescriptions. Compared to the original study, the fatal overdose model captured nearly 8% more cases in the highest risk quantile. The nonfatal overdose model identified nearly 10% fewer cases in the highest risk quantile. Notably, the top risk quantile of the fatal overdose model captured more than half of the fentanyl-related overdoses, even though the model was not specifically trained to identify overdoses linked to nonprescription opioids. Risk quantiles might be useful for identifying smaller batches of data for review to facilitate overdose surveillance.

We also present a use case for model predictions that involves aggregating prescription-level predictions to the county level. Our county-level predictions represented the proportion of individuals with high-risk prescriptions (i.e., corresponded to the highest risk quantiles of each model). County-level predictions can be ranked to identify counties that might require additional

resources for overdose prevention. When compared to observed overdose rates in 2022, our models were able to identify at least half of the 10 counties recording the highest overdose rates. This is the first study to our knowledge to aggregate prescription-level predictions to the county level. Models were designed such that predictions could be flexibly aggregated to inform overdose prevention at any level of jurisdiction (county, state, or federal).

3.4.4 Performance Varied Across Subgroups

As in the original study, the model did not perform consistently well across all age groups. This might be attributable to case imbalance and small sample sizes. Interestingly, the models did not perform as well for prescription opioids when compared to prescriptions of benzodiazepine and gabapentin respectively. One contributing factor might be changes in prescribing practices. In the original study, nearly all of the top 15 predictors for the fatal overdose model were related to either the dosing of opioids (total MMEs, average daily MMEs) or the number of overlapping and nonoverlapping prescription opioids. Since the original study period (2012-2017), safer prescribing practices have been enforced such as the co-prescribing of naloxone, potentially resulting in the prescription of lower doses of opioids and fewer overlapping prescriptions[106].

Prior studies have used PDMP data to develop models for overdose risk prediction. Oregon's PDMP data were linked to vital record and hospital discharge data in logistic regression models for identifying individuals at high risk for opioid-related mortality[8]. Maryland's PDMP data were used in multivariate logistic regression models for fatal overdose risk prediction[5,6]. However, none of these predictive models have been prospectively validated, and none have aggregated predictions to the county level. Prospective validation is critical prior

to implementation of any predictive model for opioid-related overdose, particularly given the dynamic nature of the opioid epidemic. In addition to ensuring that the models are well calibrated and able to discriminate between cases and controls, it is important that existing disparities are not exacerbated by interventions informed by model predictions.

3.4.5 Limitations

This study had several limitations. First, the validation dataset only included 9 months of data from 2022. Delays in fatal and nonfatal overdose reporting meant that 2022 was the latest full year of data available at the time of study. Only 9 months of data from 2022 could be extracted and processed, owing to server limitations. Second, the models were developed using data from one southern state, potentially limiting the generalizability of this model to other states. Demographics and other characteristics of the validation dataset are included to facilitate transportability of the model to other state health departments. Third, the tested models only consider opioid-related overdose risk. As more novel substances continue to infiltrate the drug supply, models may need updating. Predictions are also only generated at the time of a PDMP prescription. Individuals without recent PDMP prescriptions may still be at risk of opioid-related overdoses but are not captured by our model.

3.4.6 Future Work

The academic-state partnership allowed us to seamlessly leverage the expertise of both TDH and VUMC. Future work should continue to occur through this partnership. Specifically, future studies might involve understanding how opioid-related overdose risk prediction models might be implemented at state health departments. This will involve understanding socio-

technical factors impacting how a model might inform decision-making at the public health level. It will be important to consider the perspectives of public health experts who lead public health interventions and individuals who will be responsible for managing the technical components of the model. Interviews with public health experts will provide guidance on the specific interventions that may be informed by overdose risk prediction models as well as the format in which information from these models should be conveyed to users. Future work should also investigate technical details such as the frequency at which predictions should be generated, the way predictions should be aggregated to facilitate actionability, the feature engineering of additional and current variables, and questions around model updating.

3.5 Conclusion

Prospective validation was performed on fatal and nonfatal overdose risk prediction models using statewide data from Tennessee. Models were originally developed using data from 2012-2017 and were tested in this study using data from 2022. Prescription-level predictions were aggregated to the county level and compared to observed overdose rates. Models achieved similar discrimination without retraining or updating but were miscalibrated. Several of the counties with the top 10 fatal and nonfatal overdose rates were identified by the models. Additional work is needed to understand how the model might be updated to perform well across all subgroups and how predictions might be aggregated and visualized to inform public health-level interventions.

CHAPTER 4

A TREATMENT SELECTION MODEL FOR OPIOID USE DISORDER USING ELECTRONIC HEALTH RECORD AND ZIP-LEVEL DATA

4.1 Introduction

Opioid use disorder (OUD) is a chronic mental health disorder that is estimated to affect over 9.3 million (3.7%) adults in the United States[107]. OUD is diagnosed when the continued use of opioids significantly impacts daily life, resulting in symptoms such as increased tolerance of opioids, cravings for opioids, and withdrawals without opioids[108]. Medication is the first line of treatment for OUD, yet only 25% of individuals with OUD were taking medication for OUD (MOUD) in 2022[109]. Most studies have reported MOUD retention rates ranging from 20-50%, indicating that treating OUD remains a challenge[11].

In the hospital, either buprenorphine or methadone may be prescribed as MOUD to stabilize individuals at risk of overdose, per national treatment guidelines[61]. While buprenorphine and methadone have both been effective for reducing illicit drug use and overdose risk, the decision between buprenorphine and methadone is complicated by several factors[71]. Methadone is a full agonist that offers more pain control, but severe adverse events can occur if dose is increased too quickly during induction[63]. Methadone has historically been used for OUD and is covered by more insurance plans[65]. However, once patients are discharged from the hospital, initial daily doses of methadone must be administered at opioid treatment programs (OTPs). Meanwhile, buprenorphine is a partial agonist that carries a lower risk of severe adverse events and misuse[62]. Induction on buprenorphine requires patients to be in mild withdrawals first, but additional doses of buprenorphine can be obtained from any prescriber after discharge.

Selecting an MOUD for initiation therefore involves considering complex interactions among relevant clinical, social, and external factors.

Previous work has revealed that OUD treatment decisions are often led by specialists with training in areas like general and addiction psychiatry[110]. Significant differences in MOUD prescribing rates were found when specialists like addiction psychiatrists were involved compared to when they were not[12,13]. In settings where specialists are available to lead the MOUD decision and induction process, non-specialists may elect to consult their peers. However, specialists may not always be available. Further, as prescribing regulations continue to increase access to MOUDs, more non-specialists will need to initiate MOUDs[69]. One way to empower non-specialists' treatment decisions might be through a predictive model that recommends a specific MOUD for a given patient.

Previous studies have typically involved only one MOUD or have identified risk factors for treatment discontinuation. Predictive modeling has been used to predict early dropout from OUD treatment[14–16]. However, no studies to our knowledge have attempted to inform the decision between MOUDs. Treatment response has been extensively studied for major depressive disorder (MDD), which commonly co-occurs with OUD[79]. Treating MDD is challenging and typically involves trial and error[83]. Sheu et al. have been the first to use real-world, widely available electronic health record (EHR) data for predicting antidepressant treatment response[85]. Models from Sheu et al. showed good performance and generate probabilities of response for each antidepressant class. The generalizability of this modeling approach to OUD treatment selection is unknown.

In this study, we adapted the modeling framework from Sheu et al. to OUD. We developed and validated a predictive model recommending MOUD for initiation, with the goal

of empowering non-specialists to make OUD treatment decisions like specialists would. Owing to the importance of social factors in the MOUD decision-making process, we also evaluated the predictive power of including different combinations of data in the model: (1) individual-level information from the EHR only, (2) ZIP-level information from publicly available data sources only, and (3) both (combined via early and late fusion methods).

4.2 Methods

We developed and internally validated a predictive model for recommending MOUD. This study was approved by the Vanderbilt University Medical Center (VUMC) Institutional Review Board (#171323, 151156).

4.2.1 Data Sources

Models were developed using multimodal data describing individual- and ZIP-level information. Individual-level information was extracted from the VUMC Research Derivative, which contained EHR and medication data for nearly 3.4 million individuals as of October 9, 2024. EHR data were limited to the study period of January 1, 2018 to December 4, 2023. ZIP-level information was derived from public data sources managed by institutions such as the Centers for Disease Control and Prevention (CDC), United States Department of Agriculture (USDA), United States Census Bureau, University of Wisconsin, and the Tennessee Department of Health (TDH).

4.2.2 Cohort Definition

The study was limited to the index hospitalizations for individuals initiating buprenorphine or methadone in the hospital during the study period. Index hospitalizations corresponded to the first inpatient MOUD exposure during the study period. Inpatient MOUD exposures were identified using Rx-Norm ingredients. Individuals were required to have evidence of an OUD diagnosis in both billing codes and clinical notes. Specifically, individuals were required to have at least one International Classification of Disease, revision 10, Clinical Modification (ICD-10-CM) for opioid-related disorders or complications of opioid use (F11*, T40.0*, T40.1*, T40.2*, T40.3*, T40.4*, T40.6*)[12]. Individuals also had to have a clinical note mentioning OUD ("opioid use disorder," "oud," "moud," "opiate use disorder") or a chief complaint containing "heroin"[12]. MOUD initiation was determined using the total dose of MOUD administered on the first day of hospitalization (less than 40 milligrams for methadone and less than 12 milligrams for buprenorphine).

Individuals with a documented allergy to an MOUD ingredient (methadone, buprenorphine, or naloxone) or a critical methadone contraindication were excluded. Critical methadone contraindications were (1) severe liver dysfunction, defined as a Child Pugh score corresponding to Class C, and (2) QT prolongation, defined as a QT interval corrected by the Fridericia formula to a value greater than 500 milliseconds[111,112]. Individuals who did not reside in the state of Tennessee at the time of study were also excluded.

4.2.3 Outcomes

The outcome of interest was response to the initiated MOUD. Treatment response was defined as the absence of an adverse outcome within 90 days of discharge from the index

hospitalization. The 90-day period following hospitalization has been identified as a critical transitional period, during which acute stabilization is particularly important and can lay the foundation for longer term treatment planning[113]. Adverse outcomes were identified both during and within 90 days after the index hospitalization. During the index hospitalization, adverse outcomes were defined as an MOUD discontinued more than 2 days before discharge, an MOUD switch (from buprenorphine to methadone or vice versa), a hospital discharge before medically advised, or a hospital death. Within 90 days of discharge, adverse outcomes were defined as substance use-related readmissions or hospital deaths. Substance use-related readmissions were identified using ICD-10-CM codes describing nonfatal overdoses or infectious complications (bacteremia, cellulitis, discitis, endocarditis, epidural abscess, hepatitis B, hepatitis C, human immunodeficiency virus [HIV], osteomyelitis, and septic thrombophlebitis). Infectious complications were defined in consultation with a clinician.

4.2.4 Model Predictors

From the EHR, we extracted predictors recorded before and at admission using primarily structured data. Demographic factors included gender and age at admission. Insurance status at admission was modeled as a categorical variable with 6 levels: Medicaid, Medicare, Other Governmental, Private, Third Party, Unknown. Categories were not mutually exclusive, to account for individuals with multiple types of insurance. We derived several utilization measures including the number of inpatient visits, outpatient visits, and outpatient psychiatric treatment dates. The presence of any ICD-10-CM code indicating housing or economic instability was noted (ICD-10-CM: Z59*). The total numbers of benzodiazepine, stimulant, and opioid prescriptions were included.

Clinical comorbidities were also included in the model. Cardiac and respiratory comorbidities included acute asthma, heart failure, arrhythmia, hypercapnia, myocardial infarction, palpitations, paralytic ileus, and seizures. Infectious complications were bacteremia, cellulitis, discitis, endocarditis, epidural abscess, hepatitis B, hepatitis C, HIV, osteomyelitis, and septic thrombophlebitis. Psychiatric comorbidities included mood, personality, anxiety, and somatoform disorders, as well as suicidal behavior and suicidal ideation. Substance use-related comorbidities included history of substance use (alcohol, cannabis, cocaine, heroin, opioid, benzodiazepines, stimulants, fentanyl), substance use-related admission, urine drug screen at admission, number of chronic pain conditions, and history of long-term opioid use (ICD-10-CM: Z79.891). Chronic pain conditions included low back pain, rheumatoid arthritis, osteoarthritis and allied disorders, migraine, irritable bowel syndrome, endometriosis, sickle cell disease, general pain, tension headache, interstitial cystitis, vulvodynia, Ehlers-Danlos syndrome, fibromyalgia, chronic fatigue syndrome, and temporomandibular arthralgia.

Clinical comorbidities were also approximated using lab test results. Renal function was approximated using estimated glomerular filtration rate (eGFR) values. Liver function was estimated using lab values from liver function tests (alkaline phosphatase, albumin, aspartate transaminase, conjugated bilirubin, alanine transaminase, total bilirubin, and total protein). Renal and liver function tests were determined to be missing not at random, since such tests would only be administered to patients meeting a sufficient level of clinical suspicion. To allow the model to learn from this type of missingness, two variables were included for each lab test: one representing the lab value (continuous), and the other indicating whether the lab value was missing (1 if missing, 0 if not missing)[114]. Missing lab values were imputed to zero and had a nonzero value for the corresponding variable indicating missingness. Non-missing lab values

were represented by a nonzero lab value and a zeroed out missingness indicator. This imputation method enabled the model to assign separate weights to missing and non-missing lab values.

ZIP-level data were publicly available, socioeconomic indicators used by Ripperger et al. for opioid-related overdose prediction from Tennessee statewide data[9]. When information was only available at the county level, the value for the county containing most of the ZIP code was used. Factors corresponding to the most recent years were used. Namely, county-level predictors were measures of economic health (United States Census Bureau, American Community Survey - 2017), area deprivation index (University of Wisconsin – 2000), substance use coalitions in Tennessee within a 60-mile radius (TDH – 2019), number of licensed buprenorphine providers (TDH – 2020), fatal and nonfatal overdose rates (TDH – 2017), Tennessee vulnerability index (TDH – 2019), social vulnerability indices (CDC - 2018), rural-urban continuum codes (USDA – 2013), number of licensed pain clinics within a 60-mile radius (TDH – 2020), and the number of licensed methadone clinics within a 60-mile radius (TDH – 2017)[115–120]. Missing values for the numbers of pain coalitions, pain clinics, and buprenorphine providers were assumed to be zero. Missing values for vulnerability indices and area deprivation indices were mapped to the median across all Tennessee ZIP codes represented in the dataset.

4.2.5 Univariate Statistical Analysis

To compare the individuals receiving methadone to those on buprenorphine, we performed univariate statistical tests across outcomes and EHR-level predictors. For continuous variables, a t-test was used to measure the difference in means across the two groups. For categorical and binary variables, a Chi-squared test was used to determine if proportions significantly differed between the two groups.

4.2.6 Modeling Plan

The modeling plan was adapted from Sheu et al. and expanded to include testing of early and late fusion methods (**Figure 4**). In each model, the MOUD initiated was modeled as a binary variable (1 = methadone, 0 = buprenorphine), and the output was the probability of treatment response for the specific MOUD.

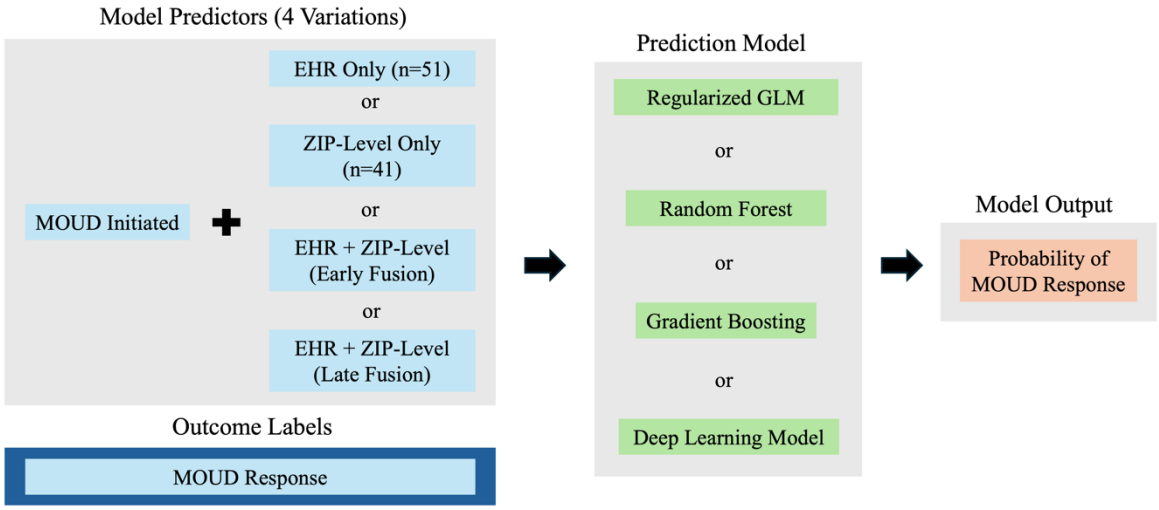
Model architectures were constructed using various combinations of model types and data configurations. Four types of models were considered: generalized linear model (GLM), random forest, gradient boosting machine, and feed-forward deep neural network (DNN). The DNN had four hidden layers (dimensions: 1024, 512, 256, 128), each followed by a dropout layer. Data configurations were EHR only, ZIP only, EHR + ZIP (early fusion), and EHR + ZIP (late fusion) (**Appendix B – Figure 11**). Early fusion was performed by combining the EHR and ZIP datasets before training the model. Late fusion was performed by training the same type of model on EHR and ZIP data separately (e.g., random forest using EHR only + random forest using ZIP-level data only), before aggregating predictions to produce one scalar output. Aggregation methods included taking the minimum, maximum, and mean of the outputs from the two models, as well as training a ridge regression model on model outputs. All combinations of models and data were evaluated.

The data were randomly split into training (75%) and test (25%) sets. Five percent of the test set was randomly sampled and held out as the validation set, which was used only for parameter tuning and calibration. Hyperparameters were tuned to maximize the area under the receiver operating characteristic (AUROC). In the GLM, alpha was set to 0.5 (elastic net) and lambda was tuned. In the random forest and gradient boosted machine models, tree depth and number of trees were tuned. In the DNN, learning rate and dropout probability were tuned.

Models were calibrated by Platt scaling, which involved applying a logistic regression model to predicted probabilities of treatment response.

Figure 4: Treatment selection model architectures adapted from Sheu et al. to include early and late fusion variations.

For early fusion models, EHR and county datasets were combined before model training. For late fusion models, the same type of model (e.g., random forest) was trained on EHR and county data separately before predictions were aggregated to produce one scalar output. Aggregation methods included taking the minimum, maximum, and mean of the outputs from the two models, as well as training a ridge regression model on predictions. All combinations of model predictors and prediction models were tested.



GLM: generalized linear model.

4.2.7 Model Performance Assessment

Model performance metrics (median and 95% confidence intervals) were calculated over 100 bootstrapped samples of the test set. Model discrimination was evaluated using AUROC, area under the precision recall curve (AUPRC), accuracy, F1, negative predictive value (NPV), positive predictive value (PPV), sensitivity, and specificity. We also calculated the Matthew's correlation coefficient (MCC), which quantifies sensitivity, specificity, precision, and negative predictive value (**Appendix B – Figure 12**)[121]. MCC values range from -1 to +1, where +1

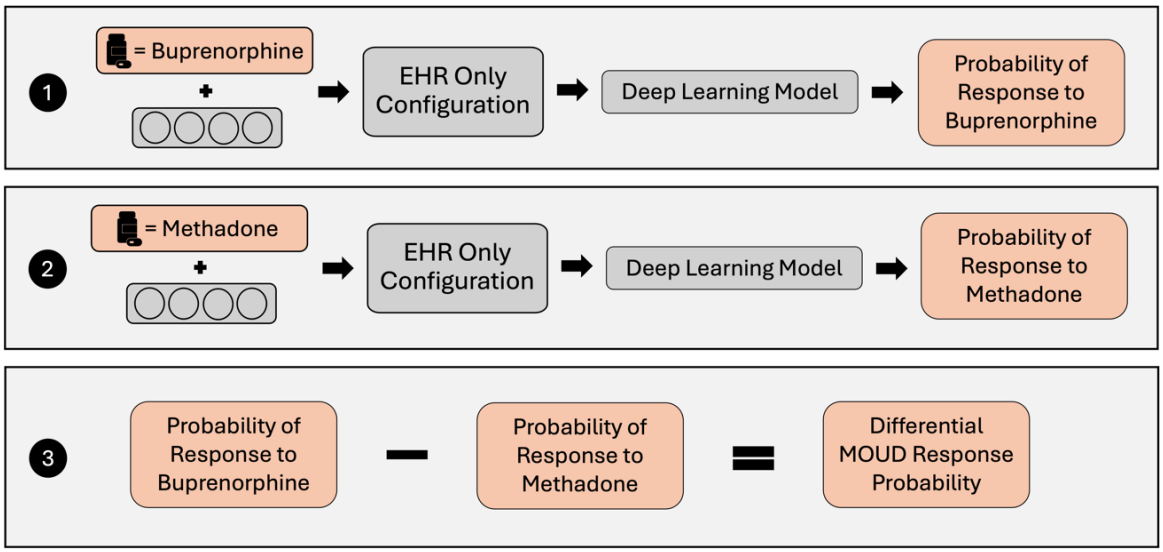
means perfect classification, 0 means random chance, and -1 means perfectly opposite classification. To calculate metrics that required a set threshold, we selected the threshold that maximized accuracy, as was done by Sheu et al[85]. Calibration was evaluated using the Brier score and Spiegelhalter z-test. The DNN model using EHR data only was used for subsequent analysis and visualization steps.

4.2.8 Predicting Treatment Response for each MOUD

For each predictor, global SHapley Additive exPlanation (SHAP) scores were generated by taking the mean absolute value of local SHAP values across the test set[122]. For each individual in the test set, we also generated personalized predictions for the likelihood of responding to each MOUD. This involved creating an expanded test set that included predictions for the alternative MOUD (i.e., methadone if buprenorphine was initiated, or vice versa) in addition to the actual MOUD initiated. The MOUD variable was toggled to either methadone or buprenorphine to obtain the respective MOUD response probability. SHAP scores were generated on the expanded test set and used to determine the specific predictors that were most influential for individualized medication recommendations. The differential probability of treatment response was calculated by taking the difference between the MOUD response probabilities for a given individual (**Figure 5**).

Figure 5: Derivation of differential MOUD response probability.

For each individual, we derived the probability of responding to buprenorphine and methadone respectively, as well as the difference between these probabilities (i.e., the differential MOUD response probability).



4.2.9 Post-Hoc Analysis

In a post-hoc analysis, we hypothesized that the differential MOUD response probability was informative for the decision between buprenorphine and methadone. To test this hypothesis, we compared the association between the differential MOUD response probability and a silver standard for the MOUD decision, which we defined as specialist-led MOUD decisions.

To create a set of silver-standard decisions, we used regular expressions to identify hospitalizations in which specialists were consulted. Regular expression patterns matched note templates typically used by addiction psychiatry, pain medicine, and general psychiatry teams. Only clinical consult notes mentioning OUD ("opioid use disorder," "oud," "moud," "opiate use disorder") were considered. We assumed that (1) if a specialist had been consulted, the specialist likely also led the MOUD decision, and that (2) the specialist was equally comfortable with initiating both MOUDs.

We tested the association between the MOUD initiated (buprenorphine vs. methadone) and the differential MOUD response probability. Specifically, we used a Wilcoxon rank-sum test to determine if the median shift in differential MOUD response probability was greater than zero when specialists decided to initiate buprenorphine vs. methadone. We repeated this analysis on the expanded test set, to also include decisions led by non-specialists.

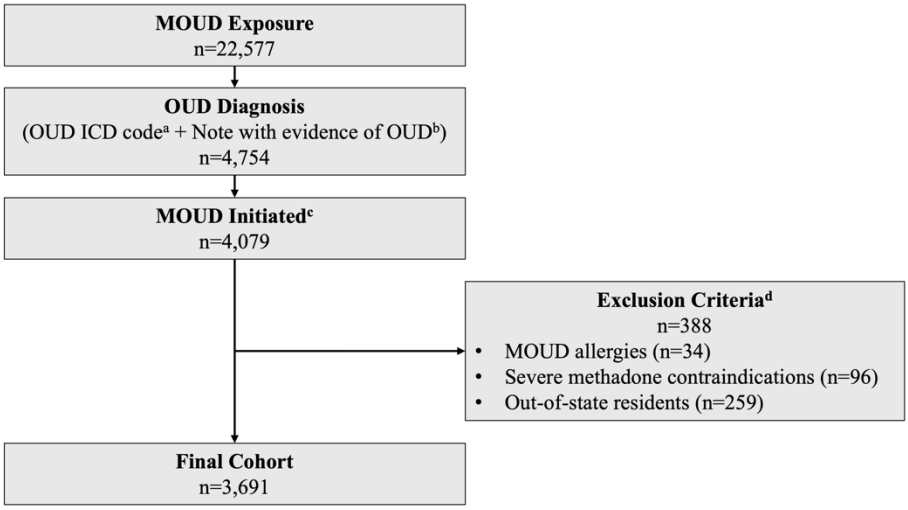
4.2.10 Experimental Setup

The DNN was trained, tuned, and tested in Python (version 3.11.5) using *PyTorch* (version 2.4.0) and *Optuna* (version 3.6.1). All other models were constructed in R (version 4.3.1) using *H2O* (version 3.44.0.3). Calibration was tested using the *rms* package (version 6.8.1) in R. Visualizations were generated using *ggplot2* (version 3.4.4) in R and *SHAP* (version 0.42.1) in Python.

4.3 Results

A total of 22,577 individuals were exposed to an MOUD in the hospital between 2018-2023 (**Figure 6**). Approximately 21% (n=4,754) of individuals exposed to an MOUD potentially had an OUD diagnosis. Most (n=4,079; 86%) individuals with OUD were given doses of MOUD corresponding to induction guidelines. A subset of individuals (n=388; 9.5%) were excluded from the cohort for allergies to at least one of the MOUDs, severe methadone contraindications, or out-of-state residency. The final cohort contained 3,691 individuals.

Figure 6: Individuals initiated on MOUD in the hospital at VUMC from 2018-2023.



a. ICD10-CM: F11*, T40.0*, T40.1*, T40.2*, T40.3*, T40.4*, T40.6*
b. Note with "opioid use disorder," "oud," "moud," or "opiate use disorder" OR "heroin" in chief complaint
c. Total dose on first day (methadone <40 mg OR buprenorphine <12 mg)
d. Applied sequentially, starting with MOUD allergies and ending with out-of-state residents

4.3.1 Comparison of Individuals on Methadone vs. Buprenorphine

Most individuals were initiated on buprenorphine (n=2,993; 81%) instead of methadone (n=698; 19%) (**Table 4**). A higher proportion of males were started on methadone than buprenorphine ($p < 0.001$). All liver and renal function tests were significantly different in either raw values or whether the measurement was taken ($p < 0.001$). More individuals with Medicaid were started on buprenorphine (32%) than methadone (24%), while more individuals with Medicare were started on methadone (24%) than buprenorphine (9.6%) (both $p < 0.001$).

Buprenorphine was more often prescribed to individuals with a history of suicidal behavior or suicidal ideation respectively (both $p < 0.001$). Individuals on buprenorphine were more likely to have more housing instability codes ($p < 0.001$)

Compared to buprenorphine, a higher proportion of individuals prescribed methadone had HIV or chronic pain conditions (both $p < 0.001$). Individuals prescribed methadone received

more opioid and benzodiazepine prescriptions compared to those prescribed buprenorphine ($p < 0.001$). The number of outpatient psychiatric treatment visits, inpatient visits, and outpatient visits were higher among individuals prescribed methadone than buprenorphine ($p < 0.001$).

Table 4: Univariate comparisons of individuals on methadone vs. buprenorphine.

Median and interquartile ranges were reported in square brackets for continuous variables. Counts and percentages were reported for binary and categorical variables. P-values were from t-tests for continuous variables and Chi-squared tests for binary and categorical variables. Race was not included in the model but is detailed below.

			Cohort (n=3,691)	Methadone (n=698)	Buprenorphine (n=2,993)	P-value[a]
Demographics	Age		36 [30, 46]	42 [33, 54]	35 [29, 44]	0
	Gender	Female	1,804 (49)	303 (43)	1,501 (50)	0.0015
		Male	1,887 (51)	395 (57)	1,492 (50)	0.0015
	Race	White	3,215 (87)	601 (86)	2614 (87)	0.42
		Black	322 (8.7)	74 (11)	248 (8.3)	0.06
		Unknown[b]	84 (2.3)	12 (1.7)	72 (2.4)	0.34
		Other[c]	49 (1.3)	9 (1.3)	40 (1.3)	1
		Multiple	21 (0.57)	2 (0.29)	19 (0.63)	0.41
	Ethnicity	Not Hispanic/Latino	2,700 (73)	492 (70)	2,208 (74)	0.086
		Hispanic/Latino	864 (23)	190 (27)	674 (23)	0.0095
		Unknown[d]	127 (3.4)	16 (2.3)	111 (3.7)	0.083
	Insurance Coverage	Medicaid	1,121 (30)	166 (24)	955 (32)	3.2e-05
		Medicare	454 (12)	166 (24)	288 (9.6)	2.1e-24
		Private	488 (13)	127 (18)	361 (12)	2.2e-05
		Other[e]	54 (1.5)	22 (3.2)	32 (1.1)	7.8e-05
		Unknown	1,581 (43)	219 (31)	1,362 (46)	1.5e-11
Infectious Diseases	Human Immunodeficiency Virus (HIV)		78 (2.1)	27 (3.9)	51 (1.7)	6e-04
	Hepatitis B		31 (0.84)	8 (1.1)	23 (0.77)	0.45
	Hepatitis C		353 (9.6)	64 (9.2)	289 (9.7)	0.75
Liver/Renal Function Labs	Estimated Glomerular Filtration Rate (eGFR)	Value[f]	99 [82, 116]	94 [74, 113]	100 [84, 117]	0
		Missing	676 (18)	78 (11)	598 (20)	8.2e-08
	Albumin	Value[f]	4 [3.6, 4.4]	3.8 [3.3, 4.2]	4.1 [3.7, 4.4]	0
		Missing	982 (27)	245 (35)	737 (25)	2.2e-08
	Alkaline Phosphatase	Value[f]	85 [69, 110.875]	90 [72.75, 121]	84 [68, 108]	0
		Missing	981 (27)	246 (35)	735 (25)	1.1e-08
	Conjugated Bilirubin	Value[f]	0.2 [0.2, 0.3]	0.2 [0.2, 0.4]	0.2 [0.2, 0.3]	3.7e-23

		Missing	1704 (46)	368 (53)	1336 (45)	0.00014
	Aspartate Transaminase	**Value**[f]	26 [18, 43]	25 [18, 39]	26 [18, 44]	1.5e-45
		Missing	959 (26)	242 (35)	717 (24)	8.2e-09
	Alanine Transaminase	**Value**[f]	23 [15, 45]	22 [14, 38.5]	23 [15, 46]	3.6e-44
		Missing	984 (27)	243 (35)	741 (25)	8.2e-08
	Total Bilirubin	**Value**[f]	0.4 [0.3, 0.7]	0.4 [0.3, 0.7]	0.4 [0.3, 0.7]	1.5e-89
		Missing	984 (27)	245 (35)	739 (25)	2.8e-08
	Total Protein	**Value**[f]	7.3 [6.8, 7.8]	7.2 [6.6, 7.8]	7.4 [6.8, 7.8]	0
		Missing	982 (27)	245 (35)	737 (25)	2.2e-08
Psychiatric Comorbidities	**Mood Disorder**		1,966 (53)	358 (51)	1,608 (54)	0.26
	Suicidal Behavior		314 (8.5)	22 (3.2)	292 (9.8)	2.8e-08
	Suicidal Ideation		870 (24)	57 (8.2)	813 (27)	3e-26
	Personality Disorder		104 (2.8)	21 (3)	83 (2.8)	0.83
	Somatoform Disorder		21 (0.57)	4 (0.57)	17 (0.57)	1
	Anxiety Disorder		1,789 (48)	380 (54)	1,409 (47)	0.00053
	Number of Chronic Pain Conditions[f]		1 [1, 2]	2 [1, 3]	1 [1, 2]	1.6e-261
Substance Use-Related	**Alcohol Use**		785 (21)	107 (15)	678 (23)	2.6e-05
	Cannabis Use		552 (15)	83 (12)	469 (16)	0.014
	Cocaine Use		476 (13)	76 (11)	400 (13)	0.09
	Heroin Use		73 (2)	5 (0.72)	68 (2.3)	0.012
	Opioid Use		3,514 (95)	661 (95)	2,853 (95)	0.55
	Sedative Use		542 (15)	75 (11)	467 (16)	0.0013
	Stimulant Use		931 (25)	159 (23)	772 (26)	0.11
	Fentanyl Use		909 (25)	191 (27)	718 (24)	0.07
	Urine Drug Screen		2,002 (54)	127 (18)	1,875 (63)	1.3e-99
	Number of Previous Overdoses[f]		1 [1, 2]	1 [1, 2]	1 [1, 2]	5.1e-06
Medications	**Number of Opioid Prescriptions**[f]		6 [2, 15]	13 [6, 47.25]	4 [2, 11]	7.7e-52
	Number of Benzodiazepine Prescriptions[f]		1 [1, 2]	1 [1, 3]	1 [1, 2]	9.7e-27
	Number of Stimulant Prescriptions[f]		8 [3, 29.25]	63 [15, 114.5]	8 [2.5, 14.5]	0.00045
	Number of Contraindications to MOUDs[f]		1 [1, 2]	1 [1, 2]	1 [1, 1]	2.2e-284
Utilization	**Number of Outpatient Psychiatric Treatment Visits**[f]		1 [1, 1]	1 [1, 1]	1 [1, 1]	1.9e-129
	Number of Inpatient Visits[f]		2 [1, 5]	3 [1, 7]	2 [1, 4]	2.2e-120
	Number of Outpatient Visits[f]		4 [1, 14]	9 [3, 32]	3 [1, 10]	1.7e-90
	Number of Housing Instability Codes[f]		1 [1, 1]	1 [1, 1]	1 [1, 1]	9.7e-34

a T-test for continuous variables, Chi-squared test for categorical and binary variables for buprenorphine vs. methadone
b Includes Unknown, Unable to Provide, Declined, Null
c Includes Asian, Alaskan/Indian, Hispanic, Middle Eastern or North African, Pacific Islander, Other, None of These
d Includes Unknown, Unable to Provide, Declined, Null
e Includes third-party liability, worker's compensation, Civilian Health and Medical Program of the Department of Veterans Affairs, and other governmental coverage
f Median, interquartile range, and statistical test only considered nonzero values.

4.3.2 Cohort Outcomes

In univariate analysis, buprenorphine was associated with better response than methadone. Across the entire cohort, 75% (n=2,767) of individuals responded to the MOUD initiated. Significantly more individuals responded to buprenorphine (n=2,441; 82% of those who received buprenorphine) than methadone (n=326; 47% of those who received methadone) ($p < 0.001$). Compared to methadone, significantly greater proportions of individuals on buprenorphine did not have adverse outcomes during the index hospitalization and within 90 days of discharge respectively ($p < 0.001$) (**Table 5**).

Table 5: Adverse outcomes for entire cohort and individuals started on each MOUD.

Treatment response was defined as the absence of an adverse outcome during the index hospitalization and within 90 days of discharge. P-values correspond to Chi-squared tests assessing the difference in outcomes for those on methadone vs. buprenorphine. Individuals with adverse outcomes both during and after the index hospitalization are counted twice.

		Cohort (n=3,691)	Methadone (n=698)	Buprenorphine (n=2,993)	Chi-Squared P-value
Index Hospitalization	MOUD Discontinued	454 (12)	279 (40)	175 (5.8)	3.3e-134
	MOUD Switched	151 (4.1)	89 (13)	62 (2.1)	4.6e-37
	Discharged Before Medically Advised	216 (5.9)	51 (7.3)	165 (5.5)	0.084
	Death	9 (0.24)	5 (0.72)	4 (0.13)	0.017
	Total	318 (8.6)	327 (47)	360 (12)	5e-100
Post Hospitalization[a]	Substance Use-Related Readmission	22 (0.6)	100 (14)	218 (7.3)	3.7e-09
	Death	687 (19)	9 (1.3)	13 (0.43)	0.018
	Total	336 (9.1)	107 (15)	229 (7.7)	3.4e-10

[a] Within 90 days of discharge from index hospitalization

In the training set (n = 2,768), a total of 2,090 (76%) individuals responded to the MOUD. In the validation set (n = 305), 222 (73%) individuals responded to the MOUD. In the test set (n = 618), 455 (74%) individuals responded to the MOUD. Across data splits, approximately the same percentage of individuals were started on buprenorphine (81%) and

responded to the MOUD (81%). The proportion of individuals responding to methadone varied slightly across data splits (48% in train, 33% in validation, 48% in test).

4.3.3 Model Performance

Models trained on only EHR data outperformed models trained on only ZIP-level information (**Table 6**). Early and late fusion models using both EHR and ZIP-level data showed performance metrics comparable to those based on only EHR data. All models using only EHR data recorded similar performance metrics. For models using only EHR data, the median values across the 100 bootstrapped samples for each metric varied slightly but did not significantly differ (AUPRC: 0.808-0.841, AUROC: 0.627-0.657, accuracy: 0.739-0.749, F1: 0.848-0.852, MCC: 0.0793-0.173, NPV: 0.556-0.69, PPV: 0.743-0.753, sensitivity: 0.983-0.998, specificity: 0.0237-0.099). See **Appendix B – Figure 13** and **Appendix B – Figure 14** for receiver operating characteristic and precision-recall curves for the DNN model based on EHR data only. Models using only EHR data were calibrated ($p > 0.05$, **Appendix B – Table 8**).

Table 6: Median and bootstrapped confidence intervals for model performance metrics in the test set.

	Data Type	AUPRC	AUC	Acc.	F1	MCC	NPV	PPV	Sens.	Spec.	Threshold
GLM	EHR Only	0.817 [0.77,0.854]	0.639 [0.588,0.68]	0.742 [0.712,0.77]	0.848 [0.822,0.87]	0.106 [0,0.276]	0.56 [0.5,1]	0.752 [0.716,0.788]	0.991 [0.9,1]	0.0398 [0,0.324]	0.327 [0.152,0.625]
	ZIP Only	0.755 [0.697,0.806]	0.528 [0.477,0.579]	0.745 [0.71,0.782]	0.851 [0.823,0.876]	0.144 [0,0.226]	0.667 [0.508,1]	0.748 [0.712,0.785]	0.991 [0.942,1]	0.0596 [0,0.2]	0.679 [0.669,0.704]
	Early Fusion: EHR + ZIP	0.81 [0.767,0.847]	0.629 [0.583,0.669]	0.744 [0.713,0.787]	0.848 [0.825,0.88]	0.176 [0.0648,0.267]	0.636 [0.525,1]	0.751 [0.719,0.79]	0.982 [0.919,1]	0.0909 [0.0136,0.281]	0.45 [0.311,0.603]
	Late Fusion: Min(EHR + ZIP)	0.809 [0.77,0.85]	0.616 [0.568,0.675]	0.744 [0.71,0.772]	0.848 [0.823,0.869]	0.133 [0,0.296]	0.573 [0.5,1]	0.752 [0.712,0.795]	0.991 [0.902,1]	0.0436 [0,0.309]	0.327 [0.152,0.625]
	Late Fusion: Max(EHR + ZIP)	0.814 [0.776,0.85]	0.629 [0.591,0.679]	0.748 [0.716,0.785]	0.849 [0.826,0.874]	0.185 [0.0934,0.266]	0.662 [0.545,1]	0.756 [0.716,0.791]	0.978 [0.937,1]	0.101 [0.0189,0.243]	0.702 [0.678,0.748]

	Late Fusion: Mean(EHR + ZIP)	0.82 [0.778,0.857]	0.643 [0.601,0.692]	0.746 [0.714,0.782]	0.85 [0.823, 0.874]	0.162 [0,0.269]	0.603 [0.512,1]	0.756 [0.721,0.794]	0.973 [0.905,1]	0.119 [0,0.339]	0.621 [0.447,0.693]
	Late Fusion: Ridge Regression(EHR + ZIP)	0.821 [0.772,0.852]	0.64 [0.596,0.686]	0.745 [0.699,0.772]	0.849 [0.821, 0.867]	0.161 [0,0.284]	0.608 [0.5,1]	0.756 [0.701,0.789]	0.975 [0.897,1]	0.113 [0,0.352]	0.523 [0.209,0.664]
Random Forest	EHR Only	0.838 [0.802,0.874]	0.654 [0.605,0.707]	0.741 [0.706,0.769]	0.849 [0.821, 0.868]	0.0974 [0,0.263]	0.69 [0.5,1]	0.743 [0.706,0.787]	0.998 [0.9,1]	0.0237 [0,0.34]	0.506 [0.441,0.683]
	ZIP Only	0.735 [0.69,0.779]	0.509 [0.463,0.57]	0.744 [0.709,0.778]	0.851 [0.828, 0.874]	0.135 [0.0274,0.201]	0.757 [0.5,1]	0.744 [0.709,0.78]	0.996 [0.977,1]	0.0385 [0.00571, 0.108]	0.585 [0.55,0.613]
	Early Fusion: EHR + ZIP	0.82 [0.782,0.864]	0.647 [0.595,0.695]	0.739 [0.707,0.769]	0.848 [0.824, 0.865]	0.162 [0,0.254]	0.587 [0.5,1]	0.748 [0.709,0.787]	0.978 [0.932,1]	0.0995 [0,0.222]	0.657 [0.568,0.675]
	Late Fusion: Min(EHR + ZIP)	0.811 [0.775,0.848]	0.615 [0.569,0.66]	0.743 [0.712,0.771]	0.85 [0.826, 0.867]	0.135 [0,0.218]	0.682 [0.522,1]	0.745 [0.713,0.774]	0.989 [0.973,1]	0.0592 [0,0.113]	0.55 [0.441,0.57]
	Late Fusion: Max(EHR + ZIP)	0.828 [0.788,0.866]	0.632 [0.577,0.693]	0.752 [0.716,0.782]	0.854 [0.828, 0.874]	0.206 [0.0897,0.286]	0.697 [0.552,1]	0.757 [0.717,0.788]	0.981 [0.961,1]	0.119 [0.0181,0.187]	0.709 [0.628,0.738]
	Late Fusion: Mean(EHR + ZIP)	0.827 [0.797,0.859]	0.631 [0.592,0.674]	0.752 [0.721,0.792]	0.855 [0.832, 0.88]	0.199 [0.137, 0.276]	0.795 [0.61, 1]	0.752 [0.721,0.791]	0.991 [0.974,1]	0.082 [0.0465,0.143]	0.633 [0.613,0.659]
	Late Fusion: Ridge Regression(EHR + ZIP)	0.832 [0.799,0.865]	0.648 [0.606,0.69]	0.751 [0.71,0.779]	0.853 [0.822, 0.873]	0.184 [0.0846,0.256]	0.694 [0.553,1]	0.753 [0.715,0.784]	0.983 [0.964,1]	0.0989 [0.0132,0.164]	0.308 [0.0772,0.347]
Gradient-Boosted Machine	EHR Only	0.841 [0.801,0.877]	0.657 [0.612,0.705]	0.739 [0.707,0.775]	0.848 [0.822, 0.869]	0.0793 [0,0.189]	0.556 [0.5,0.824]	0.745 [0.715,0.785]	0.993 [0.958,1]	0.0248 [0,0.129]	0.437 [0.368,0.551]
	ZIP Only	0.755 [0.711,0.797]	0.547 [0.502,0.59]	0.738 [0.701,0.773]	0.847 [0.821, 0.87]	0.0594 [0,0.175]	0.565 [0.5,0.75]	0.742 [0.704,0.779]	0.995 [0.971,1]	0.0225 [0,0.097]	0.709 [0.701,0.718]
	Early Fusion: EHR + ZIP	0.822 [0.792,0.866]	0.635 [0.593,0.689]	0.747 [0.721,0.777]	0.851 [0.823, 0.87]	0.179 [0.0596,0.285]	0.605 [0.505,1]	0.756 [0.726,0.791]	0.974 [0.871,1]	0.123 [0.00637, 0.372]	0.649 [0.577,0.708]
	Late Fusion: Min(EHR + ZIP)	0.815 [0.775,0.849]	0.64 [0.59,0.677]	0.739 [0.71,0.778]	0.848 [0.825, 0.873]	0.0968 [0,0.187]	0.556 [0.5,0.774]	0.75 [0.717,0.786]	0.988 [0.954,1]	0.0439 [0,0.14]	0.5 [0.368,0.539]
	Late Fusion: Max(EHR + ZIP)	0.845 [0.803,0.885]	0.665 [0.621,0.714]	0.754 [0.714,0.779]	0.853 [0.825, 0.871]	0.214 [0.105, 0.301]	0.644 [0.527,0.952]	0.76 [0.724,0.787]	0.969 [0.939,0.999]	0.147 [0.0344,0.219]	0.735 [0.718,0.745]
	Late Fusion: Mean(EHR + ZIP)	0.836 [0.798,0.872]	0.658 [0.615,0.703]	0.736 [0.702,0.773]	0.846 [0.823, 0.871]	0.0859 [0,0.224]	0.576 [0.5,1]	0.741 [0.704,0.784]	0.991 [0.958,1]	0.0326 [0,0.156]	0.601 [0.557,0.655]
	Late Fusion: Ridge Regression(EHR + ZIP)	0.839 [0.808,0.875]	0.662 [0.612,0.705]	0.738 [0.714,0.768]	0.847 [0.814, 0.867]	0.104 [0,0.276]	0.55 [0.5,0.75]	0.747 [0.723,0.792]	0.987 [0.873,1]	0.0543 [0,0.349]	0.176 [0.0385,0.608]
DNN	EHR Only	0.808 [0.773, 0.846]	0.627 [0.577, 0.671]	0.749 [0.717, 0.775]	0.852 [0.828, 0.871]	0.173 [0.1, 0.268]	0.682 [0.534, 1.0]	0.753 [0.719, 0.778]	0.983 [0.942, 1.0]	0.099 [0.013, 0.179]	0.687 [0.638, 0.74]

	ZIP Only	0.732 [0.699, 0.775]	0.5 [0.5, 0.5]	0.732 [0.699, 0.775]	0.845 [0.823, 0.873]	0.0 [0.0, 0.0]	-	0.732 [0.699, 0.775]	1.0 [1.0, 1.0]	0.0 [0.0, 0.0]	1.0 [1.0, 1.0]
	Early Fusion: EHR + ZIP	0.735 [0.696, 0.768]	0.5 [0.5, 0.5]	0.735 [0.696, 0.768]	0.848 [0.821, 0.869]	0.0 [0.0, 0.0]	-	0.735 [0.696, 0.768]	1.0 [1.0, 1.0]	0.0 [0.0, 0.0]	1.0 [1.0, 1.0]
	Late Fusion: Min(EHR + ZIP)	0.814 [0.769, 0.853]	0.633 [0.583, 0.68]	0.749 [0.717, 0.784]	0.853 [0.831, 0.875]	0.189 [0.085, 0.266]	0.7 [0.554, 1.0]	0.754 [0.718, 0.79]	0.984 [0.967, 1.0]	0.1 [0.012, 0.165]	0.686 [0.638, 0.703]
	Late Fusion: Max(EHR + ZIP)	0.738 [0.706, 0.767]	0.5 [0.5, 0.5]	0.738 [0.706, 0.767]	0.849 [0.827, 0.868]	0.0 [0.0, 0.0]	-	0.738 [0.706, 0.767]	1.0 [1.0, 1.0]	0.0 [0.0, 0.0]	1.0 [1.0, 1.0]
	Late Fusion: Mean(EHR + ZIP)	0.818 [0.773, 0.849]	0.635 [0.592, 0.682]	0.751 [0.72, 0.773]	0.853 [0.832, 0.87]	0.176 [0.083, 0.272]	0.667 [0.54, 1.0]	0.756 [0.723, 0.778]	0.983 [0.955, 1.0]	0.092 [0.009, 0.182]	0.843 [0.819, 0.862]
	Late Fusion: Ridge Regression(EHR + ZIP)	0.807 [0.774, 0.854]	0.632 [0.583, 0.687]	0.748 [0.714, 0.777]	0.851 [0.829, 0.872]	0.179 [0.096, 0.263]	0.683 [0.553, 1.0]	0.751 [0.715, 0.779]	0.984 [0.968, 1.0]	0.097 [0.012, 0.151]	0.527 [0.454, 0.553]

GLM: generalized linear model; AUPRC: area under the precision-recall curve; AUC: area under the curve; Acc.: accuracy; MCC: Matthew's correlation coefficient; NPV: negative predictive value; PPV: positive predictive value; Sens.: sensitivity; Spec.: specificity.

4.3.4 Top Predictors of MOUD Response

The number of chronic pain conditions, age at admission, and renal and liver function labs were among the top five predictors (**Figure 7**). The numbers of inpatient visits, outpatient visits, opioid prescriptions, housing instability ICD-10-CM codes, and MOUD contraindications were also important.

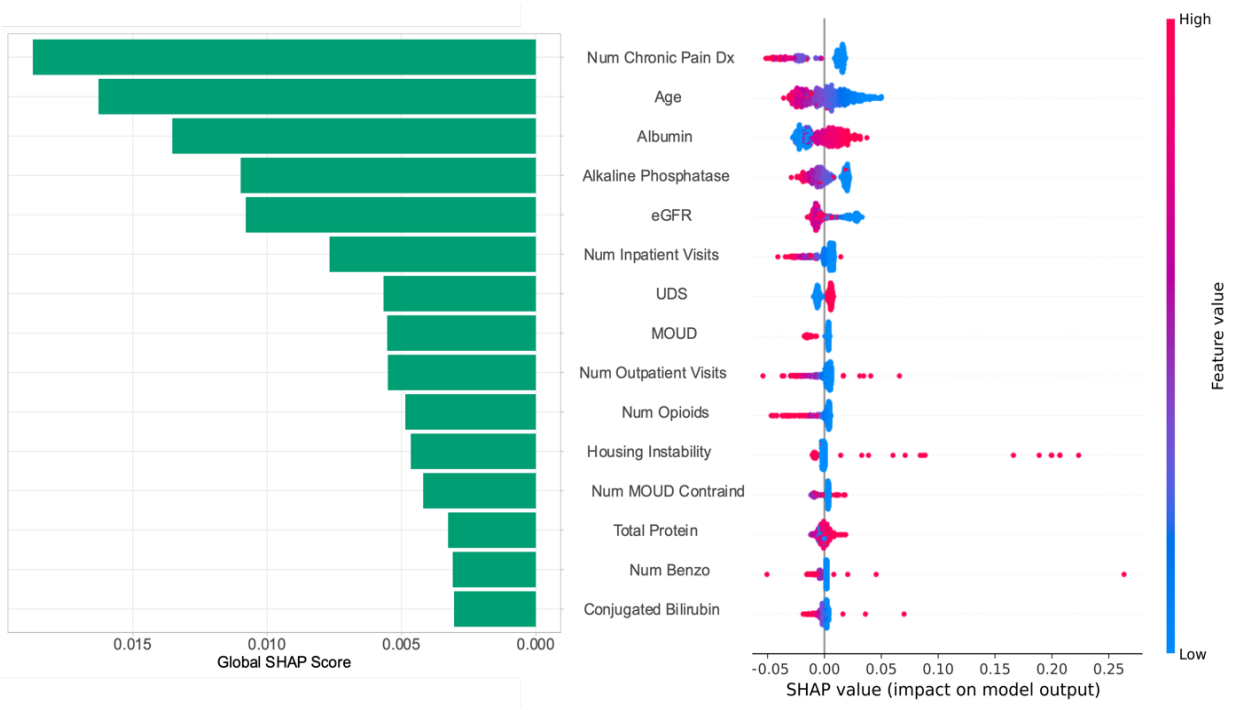
4.3.5 Probabilities of MOUD Response

In the test set (n = 618), probabilities of MOUD response for the initiated MOUD were higher when buprenorphine was prescribed ($p < 0.001$). Specialists were involved in 44% (n = 270) of MOUD decisions and initiated methadone at a significantly higher rate than non-specialists did (29% vs. 11% of hospitalizations; $p < 0.001$). In the expanded test set, in which the MOUD was toggled to generate individualized predictions for each MOUD, probabilities of MOUD response were consistently higher for buprenorphine than methadone, even when

methadone was prescribed by specialists (n = 78, 29%). The difference in predicted MOUD response (i.e., differential MOUD response) was not associated with the specific MOUD initiated across the expanded test set (**Appendix B – Figure 15**) and across decisions led by specialists only (**Figure 8**). See **Appendix B – Figure 16** for an example of how MOUD response predictions might be displayed for an individual who received buprenorphine.

Figure 7: Global SHAP scores and local SHAP values for top 15 predictors in the deep learning model.

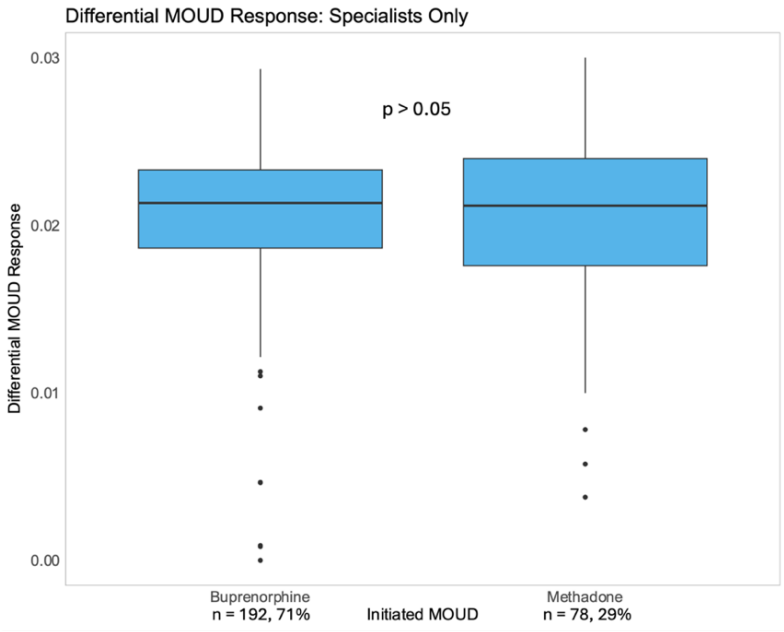
Global SHAP scores (left) are the mean absolute values of local SHAP values (right). Positive local SHAP values suggest association with treatment response.



Num Chronic Pain Dx: number of chronic pain conditions; eGFR: estimated glomerular filtration rate; Num Inpatient Visits: number of previous inpatient visits; UDS: whether urine drug screen was administered at admission; MOUD: medication for opioid use disorder; Num Outpatient Visits: number of previous outpatient visits; Num Opioids: number of previous opioid prescriptions; Housing Instability: number of ICD-10-CM codes suggesting housing instability; Num MOUD Contraind: number of MOUD contraindications; Num Benzo: number of previous benzodiazepine prescriptions

Figure 8: Differential MOUD response probabilities for individuals initiated on each MOUD, specialists only.

Medians were not significantly different by the Wilcoxon rank-sum test ($p > 0.05$).



4.4 Discussion

In this study, we developed predictive models to inform the decision between buprenorphine and methadone for OUD. Models provided probabilities of treatment response for each MOUD after considering factors describing demographics, clinical comorbidities, drug exposures, and utilization measures. Models using only individual-level EHR data outperformed models using only ZIP-level data. The addition of ZIP-level predictors to EHR data via early and late fusion methods respectively did not yield significant performance gains. Models using EHR data only performed similarly, recording AUROC between 0.627-0.657, AUPRC between 0.808-0.841, specificity between 0.0237-0.099, and MCC between 0.0793-0.173.

This study followed a previously established framework for treatment response models. One advantage of using EHR data is that this study did not require additional resources for data collection. We provide detailed counts for different types of OUD-related adverse outcomes occurring during the index hospitalization and within 90 days of discharge from the hospital. We also considered the combination of individual and ZIP-level information for MOUD treatment outcomes. The visualizations of MOUD response probabilities could allow clinicians to understand the specific factors that might have contributed to each prediction and provide a starting point for conversations with the patient about OUD treatment options.

4.4.1 Social Determinants of Health Not Well-Represented

The models recorded high AUPRC and low specificity, suggesting that the models were better at identifying MOUD response (positives) than adverse outcomes (negatives). This is not surprising, given the limitations of using EHR data for capturing social determinants and external factors that are known to affect MOUD treatment retention (e.g., access to transportation, employment status, MOUD preferences of external facilities that patients might be discharged to). Interestingly, adding ZIP-level information did not significantly improve the performance of EHR-based models. It is possible that individuals with OUD in this study were not well-represented by ZIP-level information. More individualized information might be necessary for predicting MOUD treatment response. For instance, perhaps an aggregated measure of employment across a ZIP code area is not as useful as an individual's employment status at the time of hospitalization. All models in this study recorded low MCCs, suggesting that additional model variables might be needed to better account for variation in the outcome. Further study is

needed to pinpoint the specific socioeconomic indicators that might influence MOUD treatment response.

4.4.2 Top Predictors Suggested Clinical Complexity

Several top global predictors of MOUD response suggest increased clinical complexity and replicate associations previously linked to MOUD discontinuation. A higher number of chronic pain diagnoses may require treatment offering greater pain control[123]. Elevated alkaline phosphatase and eGFR values are indicative of impaired liver and kidney functioning respectively. A greater number of previous inpatient visits was associated with a decreased likelihood of MOUD response. Inpatient visits within 6 months of treatment have previously been associated with buprenorphine discontinuation[74]. Another factor linked to a lower probability of MOUD response was previous opioid prescriptions. Soyka et al. found that the duration of continuous opioid use was linked to negative outcomes among individuals taking buprenorphine for OUD[124]. Housing instability was also associated with a lower likelihood of MOUD response and has previously been linked to lower retention on buprenorphine[125].

In this study, younger individuals were more likely to respond to an MOUD. This is in direct contrast to previous studies, which have reported an increased risk of OUD treatment discontinuation among younger individuals[14,73,74,125–128]. However, several of these studies involved only buprenorphine. In our MOUD cohort, individuals on methadone were significantly older and more likely to have more clinical comorbidities (chronic pain conditions and contraindications to MOUDs) than those on buprenorphine. The increased clinical complexity coupled with the added burden of taking methadone in the state of Tennessee, where only 24

opioid treatment programs have been registered, may have contributed to lower treatment response rates among older individuals[129].

4.4.3 Local Prescribing Practices Influenced Training Data

Across all individuals, the model consistently generated higher probabilities of response for buprenorphine than methadone, even in cases where methadone was prescribed by specialists. This is likely a result of the training data, which was influenced by local prescribing practices that heavily favored buprenorphine. The model may have inferred that buprenorphine was the safer drug, owing to the significantly higher proportion of individuals without adverse outcomes while on buprenorphine compared to those receiving methadone. Further, individuals prescribed methadone had more clinical comorbidities that may have contributed to higher rates of adverse outcomes. The model may not have had adequate opportunities to learn how healthier individuals might respond to methadone, and therefore underestimated the probability of treatment response for methadone. More complex sampling methods may improve model performance. Although differences in MOUD response probabilities were not significantly associated with the decision to initiate either MOUD, higher probabilities of MOUD response in general were linked to choosing buprenorphine over methadone. Additional study is needed to determine how to interpret the magnitude of relative differences in MOUD response probabilities.

4.4.4 Limitations

This study had several limitations. First, local prescribing practices heavily favored buprenorphine over methadone. This limited the sample size of individuals initiated on

methadone. More work will be needed as access to methadone increases and prescribing practices change. Second, outcome ascertainment was limited to EHR data from one institution. There was no way to account for admissions to other hospitals or to measure MOUD adherence (e.g., visits to opioid treatment programs for methadone or other institutions for buprenorphine). Third, socioeconomic indicators may not have been accurately represented. It is likely that the housing instability ICD-10-CM code is not consistently used in practice. More granular, patient-level information may better represent individuals with OUD than ZIP-level information. Future work might apply natural language processing-based methods to unstructured text from clinical notes to gather patient-level socioeconomic indicators like housing status. Fourth, this study was conducted using data from a state without expanded Medicaid. Findings may not be generalizable to other states with different regulations around MOUD access.

4.5 Conclusion

We developed a treatment selection model to predict the likelihood of an individual responding to medication for opioid use disorder, with the goal of informing the decision between methadone and buprenorphine. The model consistently predicted higher likelihoods of response to buprenorphine, potentially due to the makeup of the training data. Future work is needed to understand if different sampling methods might help the model provide more meaningful predictions, and if the differences in response likelihoods for the two medications could be informative.

CHAPTER 5

CLINICIAN PERSPECTIVES ON A PREDICTIVE MODEL FOR RECOMMENDING OPIOID USE DISORDER TREATMENT

The contents of this chapter have been accepted for publication in the American Medical Informatics Association 2024 Annual Symposium Proceedings[110].

5.1 Introduction

Predictive models have been developed for tasks ranging from disease identification to risk stratification. When such models have been made available as clinical decision support, however, they have not always been used[87]. A meta-analysis of clinical decision support system trials reported an overall weighted uptake of 34.2%[87]. Uptake has been linked to user factors (e.g., perceptions and autonomy) and system factors (e.g., workflow integration, usability, and technical integration with the electronic health record [EHR])[88]. In a pre-implementation study on a predictive model, Benda et al. found that stakeholders were more concerned about socio-technical factors such as how users interact with the system than with technical factors[92]. Attending to socio-technical factors may increase chance of adoption.

The specific socio-technical factors impacting the uptake of predictive models for treatment selection have not been well studied. The case of opioid use disorder (OUD) is particularly compelling because OUD treatment selection depends on clinical, social, and external factors. Consistent with national treatment guidelines, either methadone or buprenorphine (often combined with naloxone under the brand name Suboxone) is prescribed as medication for OUD (MOUD) in the acute care setting for initial stabilization[61]. There is relative

equivalence across many treatment outcomes in most studies (with a possible slight benefit for retention with methadone)[71]. Methadone has been used for OUD since the 1960s[63]. Due to prescribing regulations, methadone is generally only available from opioid treatment programs (OTPs), although hospitals may initiate or continue methadone in the inpatient setting. In the hospital, methadone induction is clinically challenging due to the long half-life of methadone and potential for severe adverse events if dose is titrated too rapidly. To continue methadone after discharge, patients must initially receive daily doses at OTPs, creating a significant burden. Further, methadone does not confer immediate protection from overdose while the medication is being titrated[130]. Meanwhile, buprenorphine has been used for OUD since the 2000s and may have more limited insurance coverage[65]. In the hospital, buprenorphine initiation can be uncomfortable for patients since they must be in mild withdrawal, but the goal stabilization dose is achieved rapidly, and the risk of overdose, severe adverse events, and misuse is lower than with methadone[66]. To continue buprenorphine after discharge, patients can receive prescriptions without attending an OTP. The MOUD decision thus involves clinical factors, insurance coverage, MOUD preference, and access to care.

To further complicate the MOUD decision, prescribing regulations are constantly changing. Before January 2023, clinicians were required to complete training and receive an X-Waiver before they could prescribe buprenorphine[131]. New modifications to methadone clinic regulations are anticipated in April 2024[69]. As both MOUDs become increasingly accessible, more clinicians will need to understand the nuances of each. In some large medical centers, specialists such as addiction psychiatrists may be available to lead decision-making. However, not all centers have these specialists, and even where they are available, they may not have the capacity to handle all such decisions. At Vanderbilt University Medical Center (VUMC), which

has an addiction psychiatry consultation service for many but not all patients with OUD, we observed noticeable differences in MOUD prescribing patterns when specialists were not involved. Rates of MOUD prescribing were lower when specialists were not consulted, and buprenorphine was more widely prescribed than methadone, potentially due to provider training and the availability of buprenorphine[12].

Clinical trials and observational studies have been conducted to compare options and identify factors associated with treatment retention[71]. Predictive models for OUD treatment retention and discontinuation risk consider only one MOUD[14]. No models to our knowledge inform the decision between methadone and buprenorphine for OUD. A predictive model for MOUD selection, which would recommend either buprenorphine or methadone for a given patient, could help empower non-specialists to make high-quality decisions without needing to consult with specialists.

While developing a predictive model for MOUD selection, we sought to understand the socio-technical factors that might impact uptake of the model. We focused on decisions made in the hospital, where clinicians might be able to coordinate OUD care for patients at an increased risk of overdose[132]. We aimed to answer two research questions: (1) How could prescribers use a model that provides personalized medication recommendations for OUD? (2) What are the minimum requirements for clinicians to feel comfortable using this model?

5.2 Methods

This qualitative study involved semi-structured interviews with clinicians who prescribe MOUDs in the hospital.

5.2.1 Setting

VUMC is an academic health system in Nashville, TN. VUMC has led several initiatives to increase access to OUD care in TN. In 2018, an addiction psychiatry consultation service was established with the goal of improving rates of MOUD initiation in the hospital. The grant-funded Bridge Clinic opened its doors in 2019 to help connect patients to OUD care during the critical 90-day transitional period post hospitalization[133]. The Firefly program is funded by the Centers for Medicare and Medicaid Services to support OUD care during the prenatal and postpartum periods[134].

5.2.2 A Predictive Model for MOUD Selection

This study was conducted concurrently with the development of a predictive model for MOUD selection. The predictive model was built with the intention of informing MOUD selection for patients initiating OUD treatment in the hospital. For a given patient, the model recommends either buprenorphine or methadone as the medication that is more likely to reduce the risks of both (1) MOUD utilization changes during hospitalization and (2) adverse outcomes within 90 days of discharge. MOUD utilization changes were noted when MOUDs were switched during hospitalization (changing from methadone to buprenorphine or vice versa), MOUDs were discontinued more than two days before discharge, or patients were discharged before medically advised. Adverse outcomes were defined as opioid-related readmissions and deaths. The model considers primarily structured data elements. Factors prior to admission include substance use history, psychiatric comorbidities, chronic pain diagnoses, and prior inpatient and outpatient visits. Factors at admission include insurance, census tract information, and whether the admission was linked to opioid-related billing codes.

5.2.3 Data Collection

Two authors (LT, JSA) developed the interview guide (**Table 7**), which refers to the predictive model described above. The vignettes in the guide were written with help from ChatGPT (version 3.5) and a pharmacist (TR). An addiction psychiatrist (KK), pharmacist (TR), and 2 PhD students in biomedical informatics (MG, US) reviewed the guide.

Participants were recruited via email and snowball sampling. Up to 2 coauthors (LT, RAB) conducted each interview either online or in-person. Interviews lasted from 30 to 60 minutes. Data were recorded, transcribed, de-identified, and analyzed using NVivo (version 14 - Mac). Interviews were conducted until data saturation, which was marked by the point at which no new concepts were observed for 2 interviews[135]. Interviews were analyzed sequentially, in parallel with data collection. The Vanderbilt University Institutional Review Board deemed this study exempt (# 230825).

Table 7: Layout of semi-structured interview guide.

Section	Scenario	Topics for Questions
Current practices for MOUD prescribing	Vignette #1: Introducing MOUD as a treatment option *Mike is a 45-year-old male who was admitted for an opioid overdose. He has a history of illicit opioid use that started after he was prescribed oxycodone for a back injury two years prior. This is his first hospitalization for an overdose, and he states it was because his dealer had a "new batch." He has not tried medication for OUD and is hesitant because of his back pain and concerns of withdrawal.*	Factors affecting treatment decision, topics covered with patient
	Vignette #2: Deciding which MOUD to recommend *Joe would like to start medication for OUD and heard from friends that methadone provides immediate pain relief.*	Factors affecting MOUD selec-tion, impact of patient prefe-rence, topics covered with patient
Requirements for using a predictive model for MOUD selection	Scenario: The hospital's new predictive model *Imagine that the hospital has developed a predictive model that was built with the intention of helping you make a decision between buprenorphine and methadone. The model considers information such as substance use history, housing status, and psychiatric comorbidities. For a given patient, the model recommends either buprenorphine or methadone.*	Expectations for model development and validation, factors influencing trust in model, preferences for how recommendation is shown
	Vignette #2 continued: Model recommends the other MOUD *Imagine you are leaning towards ordering [original MOUD chosen] for Joe, but the algorithm pops up a suggestion that [other MOUD] would be better for this patient.*	Factors impacting whether initial MOUD decision is revisited

5.2.4 Data Analysis

The Sittig and Singh socio-technical model has 8 dimensions that were conceptualized as interrelated elements within the complex system of healthcare for studying health information technology (HIT) interventions (**Figure 9**; see **Appendix C – Figure 17** for our conceptualization of overlapping dimensions)[22]. Using the Sittig and Singh socio-technical model, we performed a combined deductive and inductive content analysis, searching for concepts that fit into the different dimensions of the socio-technical model, while remaining open to emergent concepts that did not fit neatly into the model categories.

Figure 9: Sittig and Singh socio-technical model dimensions.

Socio-Technical Dimension	Description
Hardware and Software Computing Infrastructure	Technical aspects of the system required for its operation
Clinical Content	Information stored in the system ranging from clinical standards to patient data
Human Computer Interface	Helps the user engage with the system and interpret information from the system
People	All individuals involved in design, development, implementation, or use of system
Workflow & Communication	Collaboration among people in healthcare system
Internal Organizational Policies, Procedures, & Culture	Institutional rules and structures often developed in response to external factors
External Rules, Regulations, & Pressures	Factors outside institution that impact design, development, implementation, use, or evaluation of HIT
System Measurement & Monitoring	Assessing the effects of HIT (e.g., availability of features, use of features, impact on outcomes, unintended consequences)

HIT: health information technology.

Three researchers (LT, MG, US) coded the interviews. The first two interviews were independently coded by all three coders. After each of these initial interviews, all three coders met under the supervision of an expert in qualitative research (JSA) for four-way meetings to reach a consensus on the coding. Subsequent interviews were each independently coded by two coders (LT and either MG or US). After coding each interview, the two coders met to reach consensus. The codebook was iteratively modified as concepts emerged. Four-way meetings

were held after every 2 to 5 interviews were coded to discuss emerging themes. Emerging themes were shown to an addiction psychiatrist (KK) to ensure accurate interpretation and understanding of the clinical context.

5.3 Results

We interviewed 13 clinicians. Ten were non-specialists in providing OUD-specific care, with training in hospital medicine (n = 5), emergency medicine (n = 2), geriatrics (n = 1), infectious diseases (n = 1), and pulmonary critical care (n = 1). Two of the non-specialists also reported staffing the Bridge Clinic, where they worked alongside specialists in OUD care. Three clinicians were specialists in providing OUD-specific care, with training in general psychiatry (n = 2) and addiction psychiatry (n = 1). Three clinicians had training in biomedical informatics. All clinicians were age 30-39 (n = 8) or 40-49 (n = 5), and more than half were male (n = 9) or white (n = 9). Clinicians had between 3 to 16 years of experience caring for individuals with OUD. Nine indicated that they either rarely (n = 6) or sometimes (n = 3) initiated MOUDs in the hospital, while 4 did so often (n = 2) or always (n = 2).

5.3.1 Socio-Technical Model Dimensions

We identified 78 concepts across the 8 socio-technical model dimensions. Each dimension had between 3 (*System Measurement and Monitoring*) and 29 (*Workflow and Communication*) concepts. We expanded the existing socio-technical model dimensions to include additional concepts from the data (**Appendix C – Table 9**). First, we interpreted *Clinical Content* to include information accessible via the internet. Second, we understood *Hardware and Software*

Computing Infrastructure to include aspects related to the development and validation of predictive models. Third, we expanded *Workflow and Communication* to include the cognitive processes that occur within one person in clinical decision-making. No new model dimensions arose from the data.

5.3.1.1. People

Clinicians recognized their limited knowledge about medications for opioid use disorder

One of the most salient subthemes under *People* was that most non-specialists reported that they were not confident in their knowledge of the medication options for OUD.

- *Personally, I don't know . . . why some people start them on methadone versus Suboxone.* – MD1HospMed
- *. . . I can't really remember what the contraindications to methadone are.* – MD6EmergMed
- *. . . I don't know a whole lot about buprenorphine . . . I'm a little bit more familiar with . . . using methadone in the setting of chronic pain.* – MD3HospMed

Education on opioid use disorder treatment is mostly informal

Clinicians without specialized training in general or addiction psychiatry had limited formal training in OUD treatment and had acquired most of their knowledge informally or through continuing medical education.

- *. . . [M]ost of the time it's like osmosis from me asking [the addiction psychiatry team]. . .* – MD1HospMed

- *We get limited training in medical school . . . there's occasionally a lecture in residency . . .* – MD6EmergMed
- *. . . [Y]ou don't get any exposure to the methadone clinics, so induction of methadone either is through a palliative care training program . . . or you see it in the hospital . . .* – MD2HospMed

Clinicians had limited familiarity with AI

Clinicians had generally at least read about AI in some capacity, while often reporting limited understanding of it.

- *. . . I know it exists and that it can do some cool stuff.* – MD2Geriatrics
- *I'm having a hard time understanding how something can predict these things.* – MD12InfectDis
- Most were unaware of having interacted with AI-based models.
- *. . . [P]robably at times I don't even know it's happening, it's because it's working well.* – MD4PulmCritCare

A few clinicians recognized that they had interacted with clinical decision support but were unsure if these tools were machine learning or AI-based.

- *We do get these pop-ups [in the EMR] . . . I would imagine that's AI-driven.* – MD11GenPsych
- *. . . [W]hat I'm thinking more of is something like the Wells score . . . I don't think that's based on big data or machine learning* [referring to a deep-vein thrombosis risk score derived by assigning points to risk factors][136]. – MD9HospMed

- *. . . I don't know if there's machine learning with[in] Epic . . . [but] sometimes with the medications . . . [there is] a score [showing] if this patient is compliant or not.* – MD1HospMed

5.3.1.2. External Rules, Regulations, and Pressures

A patient's access to care can limit medication options

Every clinician mentioned the importance of considering a patient's access to care when deciding which MOUD to prescribe. Factors such as access to transportation, housing stability, and proximity to methadone clinics all influenced the MOUD decision.

- *. . . [I]f a patient lives very far away and doesn't have reliable transportation, I generally don't recommend methadone . . .* – MD10GenPsych
- *. . . [I]t might help to know what sort of work he does, if any, and does that allow him to . . . go to a clinic 5 days a week to get medication [if prescribed methadone]. . .* – MD11GenPsych

Several clinicians described monetary pressures related to insurance coverage that might limit MOUD options. Although buprenorphine may be more accessible to some patients, out-of-pocket costs may make methadone the more viable option.

- *. . . [I]t's also easier to find providers within your network who can prescribe Suboxone.* – MD5HospMed
- *Is there a place for them . . . if they don't have insurance? . . . [M]ethadone [might be] more difficult [to sustain] versus Suboxone, but then Suboxone needing insurance [could be a barrier] . . .* -MD7AddictPsych

Facility preferences can influence medication selection

Several clinicians perceived that a strong driver of MOUD selection was the preference of the facility managing care post hospitalization. Clinicians reported that facilities such as skilled nursing facilities, rehabilitation centers, and group homes enforced additional rules around MOUD, often in favor of buprenorphine over methadone.

- *Some [skilled nursing facilities and rehabs] will take Suboxone, almost none take methadone. If they're going to a group home, sober living community . . . a lot of them have a lot of restrictions. Certain housing options won't allow [any MOUD] . . . [F]ive percent of my population is actually feasible to have methadone or Suboxone.* – MD5HospMed
- *Some shelters do UDSs [or urine drug screens] and will kick you out if your UDS comes back X, Y, and Z. And some housing places still sort of have stigma against Suboxone and consider it to not be in recovery if you're on Suboxone.* – MD2Geriatrics

Societal stigma is a barrier to receiving medication for opioid use disorder

Several clinicians acknowledged that a diagnosis of OUD or an MOUD prescription could be a source of stigma and pose additional barriers to treatment.

- *That's probably one of the most frustrating things . . . it's hard to find pain clinics that will treat somebody with a substance use disorder . . .* – MD10GenPsych
- *. . . [S]ometimes a "drug-seeking" patient is actually more of a relief-seeking patient . . .* – MD11GenPsych

- *. . . [We need to help] people understand that opioid use disorder . . . is a chronic brain disease . . . analogous to other chronic diseases that we manage longitudinally . . .* – MD12InfectDis

Difficult to keep up with prescribing regulations that are constantly changing

Clinicians wanted a model that was aware of up-to-date prescribing regulations and initiation guidelines.

- *Because it is a little intimidating with Suboxone . . . the guidelines are changing. . . . it's sometimes just really hard to keep up with.* – MD8EmergMed
- *I can foresee it ultimately saying, "Hey, methadone is the appropriate choice. Here's a brief synopsis of initiation" and guidelines and what to watch for.* – MD3HospMed

5.3.1.3. Hardware and Software

Model should consider social determinants of health and facility preferences

Several clinicians suggested that they would trust a model more if it factored in information about social determinants of health and facility preferences.

- *My concerns are if the algorithm can identify a prescriber out of the hospital, take into account insurance status, take into account post-acute care destination, and housing status, not just the housing status they come in on but the one that they're going to go to.* – MD5HospMed

Model should be validated within a sufficiently large and similar patient population

Clinicians generally wanted a validated model. For some, this meant a clinical trial evaluating the impact of the tool on relevant outcomes such as rates of hospital readmissions, MOUD prescribing, and remission.

- *. . . I would hope that whoever developed it had tried to put this in a clinic . . . –* MD12InfectDis

A sufficiently large dataset was important for many clinicians.

- *I'm always a little bit skeptical of the new technologies . . . I really would like to see [it] developed and then tested and sort of a beefy data set assuring that it's a good intervention.* – MD12InfectDis
- *I'm not gonna be scrutinizing it, I am not a researcher . . . But I do wanna make sure that they've tested it on more than, like, a handful of people . . . –* MD10GenPsych

Several clinicians wanted to make sure the model had been tested on a population similar to the one they were caring for, while a few wanted to ensure that the model also performed well across all subgroups.

5.3.1.4. Clinical Content

Concern about quality of EHR data for important model variables

Many clinicians were concerned about the completeness of data on substance use history, medication reconciliation, and housing status. These data were likely to be outdated or undocumented in the EHR. Additionally, several clinicians were skeptical that a predictive model could take into consideration any insight gained from motivational interviews, which they believed were critical for increasing OUD treatment success.

- *So, you do have to have the data of the patient and the conversation and their goals for their care, which I'm not sure can be put in a model . . .* – MD2Geriatrics
- *. . . the humanistic piece, I don't know how it can get that part . . .* – MD7AddictPsych

Strong preference for evidence-based information

When asked about information they would need to know to feel comfortable using the predictive model, all clinicians wanted evidence-based information. In addition to testing and validation, the model should be based on factors identified in literature.

- *. . . [I]f there is established science . . . then I'm going to be thrilled to take that into account and augment my practice in whatever way I can.* – MD6EmergMed
- *. . . [I]f the interruptive alert had "The other medication might be better and here's why" . . . I [would] go read more about it. And if I think "That's – oh, yeah, that's fair," then I change my practice.* – MD2Geriatrics

Disinterest towards including pharmacogenomic information in model

When asked about the importance of including genetic information in the predictive model, clinicians were not convinced that genetic data would be attainable for this patient population or useful for MOUD selection.

- *But there's so much more that goes into [OUD treatment] . . . I don't think we're there, yet.* – MD10GenPsych

5.3.1.5. System Measurement and Monitoring

Model should prioritize medication safety

Clinicians were less tolerant of model error when patient safety was the unintended consequence.

- *If the algorithm missed something big, like a prolonged QTc with methadone, that's a different level . . .* – MD2Geriatrics
- *. . . I would want to be very sure with things like methadone that some of the other history has been taken into account before I were to give it.* – MD8EmergMed

5.3.1.6. Human Computer Interface

Decision support tool should automatically generate recommendations within the EHR

Clinicians preferred a model that would automatically run from within the EHR.

- *. . . [If it tells me] in the EHR, that the recommendation is X or Y . . . I would be much more inclined to use it.* – MD9HospMed
- *. . . [U]sage would be a lot lower the more steps that I have to go to interact with it.* – MD8EmergMed

Clinicians wanted to understand how the model arrived at its recommendation(s)

Several clinicians wanted to understand the rationale for MOUD recommendations. Some were interested in explanations of the model's rationale or understanding the relative projected risk or benefit for each medication.

- *. . . [I]t's always going to give me a little bit of the heebie-jeebies when something makes a recommendation to me and can't tell me why.* – MD6EmergMed

- *So, I don't know if . . . it is able to come back and say, "I would choose buprenorphine for this one particular reason and not methadone," and it could like, tell me the pain point, so to speak.* – MD7AddictPsych
- *If it's a pretty slight difference . . . I would probably defer quite a bit more to patient preference, whereas if it were [a larger difference], then I would give stronger weight to the model.* – MD9HospMed

Decision support as a vehicle for clinician and patient education

Clinicians were open to learning from the model.

- *. . . [I]t would be interesting to no matter what see . . . why the model disregarded the other side [of the decision tree] . . . to not only understand and confirm its accuracy, but then it's a teaching tool too . . .* – MD2Geriatrics
- *. . . [I]t would be a good hard stop to consider . . . what's going into my decision, and then . . . take whatever the model is telling me into consideration.* – MD11GenPsych

Some clinicians envisioned the decision support tool shaping conversations with patients.

- *. . . I do always appreciate it if there's a little blurb describing why. . . . [B]ecause that might frame, again, the conversation that we have with the patient at the bedside later.* – MD12InfectDis
- *. . . I can take something and give it to the patient and say, "Look, this is what our model printed out. This is 100 little people and you're the little red person here."* – MD6EmergMed

5.3.1.7. Internal Organizational Policies and Procedures

Clinicians were more comfortable initiating buprenorphine

Non-specialists primarily initiated buprenorphine in the hospital and had limited experience with methadone.

- *. . . I'm a Suboxone girl myself . . . I probably would not lean towards methadone.* – MD2Geriatrics
- *I've never initiated a patient on methadone . . .* – MD9HospMed
- *. . . [W]e generally only do buprenorphine. I would not feel comfortable initiating methadone without [expert consultation] . . .* – MD8EmergMed
- *. . . I just do what my peers do, put people on buprenorphine and send them to clinic.* – MD6EmergMed

Barriers to prescribing methadone included the increased risk of QT prolongation and the federal requirement for methadone to be administered at specific clinics.

- *. . . [B]uprenorphine in general is so much safer and . . . [easier] to continue . . .* – MD2Geriatrics
- *But methadone has a lot of regulations around it. And because of the concerns for cardiac risk, you have to really make sure you have all your ducks in a row.* – MD4CritCare

Addiction psychiatry team often oversaw initiation of medication for opioid use disorder

Non-specialists were grateful to have access to an addiction psychiatry team and typically called in addiction psychiatry colleagues to help make decisions around MOUDs.

- *. . . [T]hankfully because we have addiction psychiatry I don't have to . . . figure out how to initially dose this and monitor this.* – MD3HospMed
- *I suppose I could in a pinch [figure out how to initiate methadone], but in my context . . . I would consult addiction psychiatry and get their guidance on that.* – MD9HospMed

Administrative oversight would foster trust in the model

Several non-specialists indicated that they would trust a predictive model if its implementation was overseen by and endorsed by the addiction psychiatry team and hospital administration.

- *. . . [I]f [addiction specialists] said "We trust this thing. . . . " then I'm in. I'm all in.* – MD6EmergMed
- *. . . I'm not sitting here like wondering, "Oh, who put this together?" "How was this developed?" I'm just like, "These are the experts, this is what they're telling me to do." So, I would make the assumption that it's been vetted quite rigorously.* – MD3HospMed

5.3.1.8. Workflow and Communication

Patient readiness and preference are important for treatment success

Clinicians emphasized the importance of understanding patient readiness and adjusting the care plan appropriately.

- *. . .[I]f the patient doesn't want it . . . it's like a battle.* – MD1HospMed

- *. . . [I]t's their decision . . . sometimes they say, "No, thanks." And then we . . . make them stable, send them with Narcan, and hope that maybe the next time we can add another brick to the road.* – MD7AddictPsych

For those who are ready to receive MOUD, factors such as living situation, prior MOUD attempts, and MOUD preference are important to consider before informing the patient about options and recommending an MOUD.

- *. . . [A] lot of patients . . . come in and say, " . . . I had an allergy to [Suboxone]. So, I'm not gonna touch it again with a 10-foot pole." And sometimes what really happened was . . . it caused a precipitated withdrawal . . .* – MD11GenPsych
- *But if they really want to do methadone and we make them do buprenorphine, then that might be the reason why they're failing therapy. We didn't ask the patient what they wanted to do.* – MD12InfectDis

Opioid use disorder care calls for complex considerations

Clinicians described a complex clinical reasoning process that involved addressing pain and withdrawal concerns, prioritizing medication safety, considering substance use history, and promoting harm reduction practices. These clinical considerations must be weighed alongside patient preferences and factors described in *External Rules, Regulations, and Pressures*.

- *They might want methadone all day long, but if they don't have a car or a way of getting to a clinic an hour away every day, that's not setting them up for success.* – MD7AddictPsych

Opioid use disorder care involves several different roles

One of the challenges to providing OUD care is coordinating care for patients. In the hospital, clinicians in the emergency department and intensive care unit saw their role as bridging the gap in care before their patient could be connected to longer term care options.

- *. . . [If] this thing popped up at me for every single overdose patient . . . I would just go, no, get the **** out of here. . . . [T]heir brain is toast from their cardiac arrest from all of this, I don't care [about starting them on OUD treatment].* – MD4CritCare
- *I'm not making long-term decisions for these people; I'm trying to keep them comfortable for a couple [of] days until they can see an actual specialist . . .* – MD6EmergMed

Post hospitalization, care coordination involves considering factors such as those previously discussed under *External Rules, Regulations, and Pressures*. Several clinicians mentioned the importance of social workers, recovery coaches, and support networks after patients leave the hospital.

- *We could bring in recovery coaches, have more of a peer-to-peer conversation, talk to their family, try to get some more collateral buy-in . . .* – MD7AddictPsych

One clinician noted an unclear understanding of OUD care responsibility.

- *So there's inertia . . . who owns it? And then if they're seeing a psychiatrist or a pain clinic . . . do I own that or do they or does someone else own it?* – MD4CritCare

The model should complement existing workflows

Most clinicians wanted to see the model recommendation early in their workflow, before talking to the patient. As mentioned earlier in *Human Computer Interface*, clinicians could see

the model informing their conversations with patients. As mentioned in *Clinical Content*, clinicians did not believe this model could replace the conversations.

- *I still think walking into the room with 2 effective treatments and asking patients what they want to do would be the same. But having more information to talk to a patient about is always helpful.* – MD12InfectDis

Clinicians wanted to understand the goals and limitations of the model, so they would be able to appropriately weigh model recommendations against other factors in their decision-making process.

- *. . . [I]s there some guideline that it's built off of or is this expert recommendation?* – MD4CritCare
- *. . . [I]t would be up to the provider to take into account the psychosocial aspects of use and their decision.* – MD10GenPsych
- *I would just like to know [the characteristics of the training data] so that when I'm sitting down and using the model I can say, "Hey, maybe not go with this recommendation" or maybe I'll take it less into consideration.* – MD13HospMed

Clinicians were concerned about potential unintended consequences if clinicians followed recommendations from the model without considering other factors that might influence MOUD selection.

- *If people use that as their sole decision-maker, just like with any test, it could lead them down the wrong path. . .* – MD10GenPsych
- *. . . [T]he model would almost have to be an educational tool, really making sure providers who chose the methadone pathway understood all of the ramifications . . .* – MD2Geriatrics

- *I would use it as an additional [source of information]– because we always say, "Hey, you've gotta treat the patient in front of you."* – MD11GenPsych

5.4 Discussion

Through semi-structured interviews, we found that a predictive model for MOUD selection could address knowledge gaps for non-specialists, who reported having limited knowledge about MOUD. External factors such as access to care, facility preferences, societal stigma, and prescribing regulations all impacted MOUD selection. In addition to clinical comorbidities, the model should consider social determinants of health, facility preferences, and evidence-based information. Before deployment, a model should be tested within a large enough dataset and similar patient population. Clinicians wanted to understand how a model arrived at its recommendation. Non-specialists relied on the addiction psychiatry team to lead MOUD decisions and were willing to trust a model that was endorsed by their specialist peers. OUD care requires considering clinical, social, and external factors, as well as engaging the patient and coordinating care after hospitalization. The goals and limitations of the model should be apparent, and an educational component should accompany the tool, so users can understand how to put recommendations into practice. We also identified key takeaways that span multiple dimensions of the Sittig and Singh socio-technical model.

5.4.1 Non-Specialists Trusted Specialists and the Institution

Non-specialists often trusted and relied on their specialist peers to initiate MOUD in the hospital. Most non-specialists were aware of the risks of cardiac-related adverse events associated with methadone, and were more comfortable prescribing buprenorphine owing to

training, accessibility, and medication safety. When asked about a predictive model for MOUD, several non-specialists cited expert oversight and endorsement as sufficient criteria for uptake. Other studies have also noted the important role of champions for the successful implementation of informatics interventions[137]. In our study, the institution was believed to have processes in place to screen and evaluate decision support tools before they were deployed. This institutional trust underscores the importance of ensuring that the institutional infrastructure does in fact exist to support the involvement of clinical experts in the design, development, implementation, and evaluation of HIT interventions.

5.4.2 Clinicians Expected the Predictive Model to Think Like a Clinician

Clinicians expected the predictive model to make decisions as a clinician would, including considering social and clinical factors (patient readiness, social supports, access to care, and comorbidities). Recommendations should be evidence-based, prioritize safety, and adhere to up-to-date prescribing regulations. Clinicians wanted explanations, particularly if the recommendation differed from the clinician's judgment. Large language models are capable of emulating this level of explanation and might be used to generate such explanations in future work[138]. The ultimate treatment decision should be up to the clinician. Other studies have also recognized the importance of clinician autonomy[88,139].

5.4.3 Clinical Reasoning for MOUD Decision Requires Considering Information Not Well-Captured by the EHR

MOUD decisions require factors that are often changing and difficult to capture in the EHR. Access to care involves understanding social determinants of health such as employment

status, transportation needs, and housing status, all of which are subject to data quality challenges in the EHR[140]. The patient's MOUD preference may be influenced by prior MOUD attempts or interactions with peers. Preferences may change after the clinician educates the patient about both MOUDs. This means that patient preference is only knowable through conversations with the patient, which may not be documented in the EHR. Interestingly, many clinicians reported that facility preferences often limited MOUD options, although the American Disabilities Act prohibits denying patients with OUD access to health services[141]. Facility preferences are not readily available in the EHR. These limitations to data availability suggest that a predictive model for MOUD will likely be unable to reproduce the clinical reasoning process.

5.4.4 Predictive Models for OUD Treatment Should Be Designed to Foster Appropriate Trust

These modeling challenges highlight the need for MOUD predictive models to be designed with the goal of fostering appropriate trust. Appropriate trust can be facilitated by allowing users to consider the purpose of the tool, its process for generating recommendations, and its performance in different contexts[142]. Clinicians should be made aware of the limitations of the model and should understand when to interpret recommendations in the context of additional information from the EHR or the patient. For instance, a clinician might choose to follow the model recommendation when one MOUD is contraindicated. In cases where the MOUD decision is less clear-cut, the clinician might opt to use information from the model to inform the conversation with the patient. Future work might study how visualizations or different sorts of explanations might impact clinician behavior. As more predictive models are deployed,

it is important that model developers, clinicians, and health IT continue to collaborate to ensure tools are useful and performing as expected[143].

5.4.5 Limitations

Several factors limit the generalizability of our findings. The setting had an addiction psychiatry consultation service and served a largely insured population. The study took place within 1 of 10 Medicaid non-expansion states[144]. We sought to reach data saturation across all specialties, rather than within each specialty, limiting our ability to extract themes from each specialty. This study was limited to a predictive model for MOUD and may not generalize to other types of models. Finally, we asked about a hypothetical predictive model still in development and used hypothetical cases; responses might differ with real predictions about actual patients.

5.5 Conclusion

Many predictive models have been developed in healthcare, but only a handful of those deployed as decision support have been widely adopted. The uptake of HIT depends on socio-technical factors such as those outlined by Sittig and Singh. We used the Sittig and Singh socio-technical model to identify factors that might impact the uptake of a predictive model for MOUD selection. Through 13 semi-structured interviews of clinicians who prescribe MOUD in the hospital, we identified key takeaways that might inform the implementation of predictive models for MOUD. Non-specialists would trust a tool that was endorsed by experts and that appeared to follow clinical reasoning. However, they also wanted to consider factors not well captured by the EHR, such as housing status, access to transportation, and facility preferences. Therefore,

predictive models for MOUD must be designed to foster appropriate trust, which accounts for the tool's purpose, process, limitations, and performance in different contexts.

CHAPTER 6

CONCLUSION

In this dissertation, we have applied predictive modeling to inform opioid-related overdose prevention at the public health and individual levels. In this chapter of the dissertation, we will discuss the specific findings from previous chapters and the general implications of these findings for the domain of opioids and the field of biomedical informatics.

6.1 Innovation

At the public health level, we prospectively validated predictive models that were previously developed by Ripperger et al. for opioid-related overdose risk[9]. The models were initially developed using data from 2012-2017. Compared to the original model development study, the models recorded similar discrimination on data from 2022 without additional retraining or updating. Models remained miscalibrated after applying an off-the-shelf calibration method, suggesting that more complex calibration techniques may be needed. Shifts in the prevalences of cases and causal factors (e.g., prescribing regulations, drug supply) over time likely contributed to the observed miscalibration[105]. Future work might involve measuring the shifts in data between the original and prospective validation studies, as well as characterizing the groups for which the model over- and under-predicted risk. This was the first study to our knowledge to test the external validity of an overdose risk prediction model using statewide prescription data. It is notable that the model remained valid on newer data, despite the increase in overdose deaths driven by synthetic opioids and other substances since 2012[20]. Socioeconomic factors derived from publicly available data sources also did not need to be updated for the

model to achieve this level of performance. Additionally, we demonstrated how model predictions could be divided into risk quantiles to identify high-risk prescriptions. Using these risk quantiles, we presented one potential method for aggregating prescription-level predictions to generate county-level predictions. We showed that our approach, which involved calculating the proportions of high-risk prescriptions in each county, reproduced several trends in observed overdose rates from the same year. We emphasize that our models were developed to allow for prescription-level predictions to inform public health interventions at different jurisdiction levels (county, state, federal).

At the patient level, we developed treatment selection models for opioid use disorder (OUD), with the goal of facilitating the initiation of medication for OUD (MOUD) in the hospital. For a given patient, the model considers primarily clinical factors from the EHR and provides a likelihood of response to each MOUD. MOUD response was defined as the absence of an adverse outcome during the index hospitalization or within 90 days of discharge from the hospital. Models that were trained on exclusively EHR data outperformed models that used ZIP-level information only, and combining the two types of data yielded no significant performance gains. It is possible that individuals with OUD are not accurately represented by ZIP-level information, and that more granular patient-level information may be more predictive of MOUD response. The models had low specificity, potentially because we were unable to account for factors related to an individual's access to care or MOUD preference. Access to care and patient MOUD preference were important factors impacting retention and response, according to the clinician interviews that we conducted concurrently with model development. We are the first to our knowledge to use predictive modeling for MOUD treatment selection.

While developing the model for OUD treatment selection, we interviewed clinicians to understand factors that might influence whether the model would be used in practice. We found that clinicians tended to trust their peers and the institution. Non-specialists typically deferred to their more specialized peers to lead OUD care decisions. Many clinicians believed that decision support tools deployed in the EHR had been thoroughly vetted by the institution. Clinicians wanted the model to think like them—they wanted to be able to trace the clinical reasoning process followed by the model to assess the validity of its suggestions. Several factors in the clinical reasoning process for MOUD choice specifically involve information that is not well documented in the EHR, such as access to transportation, employment status, and the MOUD preferences of patients and external facilities. The treatment selection model could thus inform conversations with the patient, during which clinicians might gather additional information that could impact MOUD choice.

6.2 Implications for the Domain of Opioids

Our work with overdose risk prediction models has implications for similar models using statewide data. Even in the dynamic context of the opioid epidemic, overdose prediction models may not require constant retraining and updating to achieve similar model discrimination. It is particularly important to understand the frequency at which such models need to be updated, because state health departments may need to rely on external contractors or collaborators for modeling expertise and more robust servers. Model predictions may also support public health interventions that are resource-intensive, such as the dispatchment of additional naloxone shipments to certain jurisdictions. Further study is needed to determine the appropriate method

for monitoring overdose prediction models, as well as the frequency at which such models should be updated.

An overdose risk prediction model that works on data from one state may not perform as well on data from another state. Several factors pose challenges to the transportability of such models. Shifts in prescribing regulations and local drug supply may induce differences in underlying causal factors. Site-specific processes such as the reporting practices of local jurisdictions may result in differential reporting delays and inconsistent case definitions. An overdose risk prediction model based on Medicaid data has been validated across state lines, but the same has not yet been done for a model based on prescription drug monitoring program data[145]. Moving forward, it will be important to understand if overdose risk prediction models are generalizable to other states. Additional work is needed to determine the information needs of public health experts who would use such models, as well as the specific public health interventions that the model could inform.

Our work with OUD treatment selection models demonstrated the importance of including social determinants of health in models for OUD care decisions. Factors such as access to transportation and employment status are all critically important to consider and may impact whether an individual is able to continue taking MOUDs as prescribed. This was evident from our clinician interviews, where social determinants of health were a recurring theme. Although ZIP-level predictors did not significantly improve model performance in our work, it is possible that patient-level social determinants of health could better represent individuals with OUD and thus improve model performance. Future work should involve systematically collecting and studying social determinants of health at the patient level.

6.3 Implications for the Field of Biomedical Informatics

The field of biomedical informatics has benefited tremendously from the collaboration of individuals with different areas of expertise. For example, partnerships among clinicians, health information technology experts, and EHR vendors are important for creating clinical decision support tools that are useful at the point of care. Similarly, groups like the electronic medical records and genomics (eMERGE) consortium have enabled investigators to share and validate phenotyping algorithms across different institutions[146]. This dissertation highlights the importance of a type of collaboration that is less common: the partnership between an academic institution and a state health department. Academic-state partnerships have the potential to catalyze innovation in public health informatics. While public health experts at the state health department have extensive knowledge about topics like overdose reporting and surveillance, academic institutions may have access to more advanced servers that can be used to develop more advanced, AI-based models. Together, we can work towards developing modernized systems and solutions that are useful and effective for timely prevention and intervention at the public health level.

Along a similar vein, this work highlights the importance of engaging end users like clinicians early in the model development process. This is important for understanding how decisions are made at the point of care. Engaging clinicians early could reveal factors that may impact whether the model would be used if implemented. For example, the clinical reasoning process for a particular care decision may hinge on a specific factor. Accounting for this factor might then be critical for fostering trust in the model. By working closely with clinicians, we might also be able to better adapt to the specific information needs of each user. Whereas one individual may want to know the ins and outs of the predictive model, another individual may be

content with a quick walkthrough of how to use the model. Yet another individual may just ask to read a publication showing that the model has been validated. The heterogeneity in user needs is worth noting, particularly when considering how AI-based tools should be designed to foster trust.

Trust in AI is a topic that warrants additional study. This work supports the notion that AI-based tools should be developed and designed with the goal of fostering appropriate trust, which considers the purpose of the tool, its process for generating recommendations, and its performance in different contexts[142]. To foster appropriate trust, clinical decision support systems must be intentionally designed with and for clinicians. We need to understand factors like the role of clinician expertise in decision-making and how tools should be personalized to clinicians with different training. Moreover, systems that monitor AI algorithms will be necessary. Careful oversight of AI algorithms will become increasingly important as more AI-based tools are deployed. It is vital that models are monitored, updated, and decommissioned as needed over time to ensure that existing systems are supporting patient safety and not exacerbating existing disparities. The potential for AI-based models to exacerbate existing health disparities calls for safeguards to be established around the development, validation, implementation, use, and maintenance of these models[143,147].

6.4 Conclusion

This dissertation applied predictive modeling to (1) predict overdose risk, with the goal of supporting safer opioid prescribing at the public health level, and (2) to guide treatment selection for OUD, with the goal of increasing the initiation of MOUD at the individual level. Taken together, this work supports the use of predictive modeling for overdose prevention.

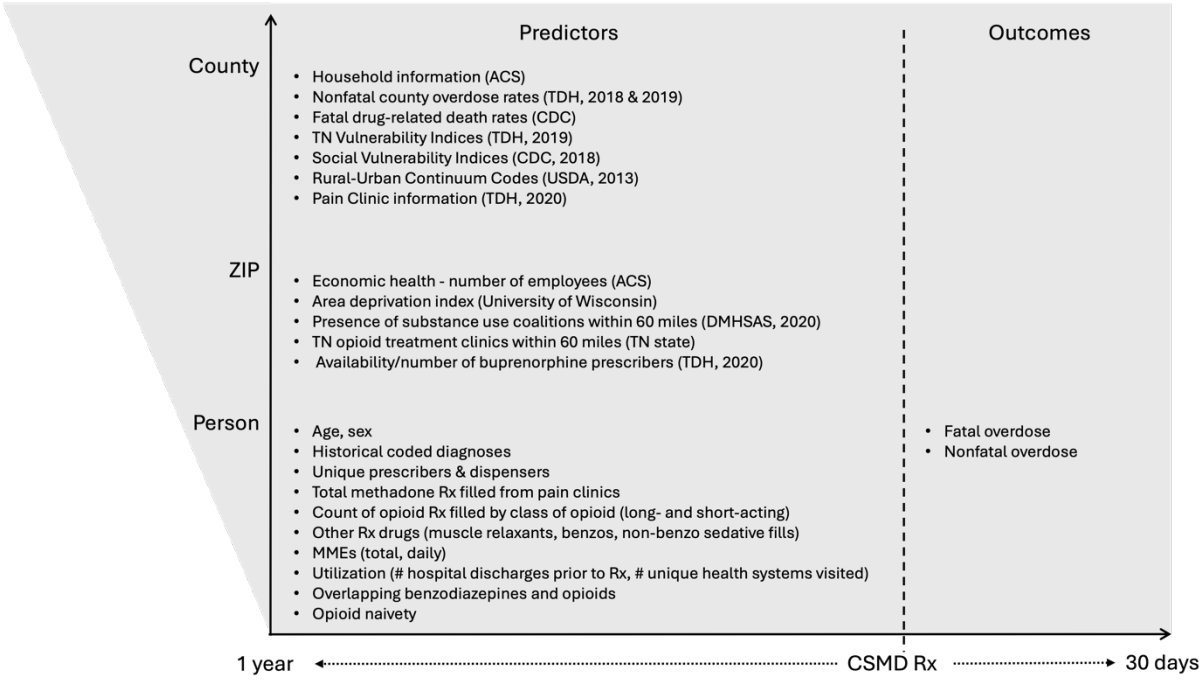
APPENDICES

APPENDIX A

Predictors for Overdose Risk Prediction Models

This appendix contains additional information on the predictors used in the fatal and nonfatal overdose risk prediction models described in Chapter 3.

Figure 10: Predictors at the prescription, ZIP, and county levels for the fatal and nonfatal opioid-related overdose risk prediction models.



Rx: prescription; ACS: American Community Survey; TDH: Tennessee Department of Health; CDC: Centers for Disease Control and Prevention; USDA: United States Department of Agriculture; DMHSAS: Department of Mental Health and Substance Abuse Services; TN: Tennessee; CSMD: Controlled Substance Monitoring Database (Tennessee's Prescription Drug Monitoring Program).

APPENDIX B

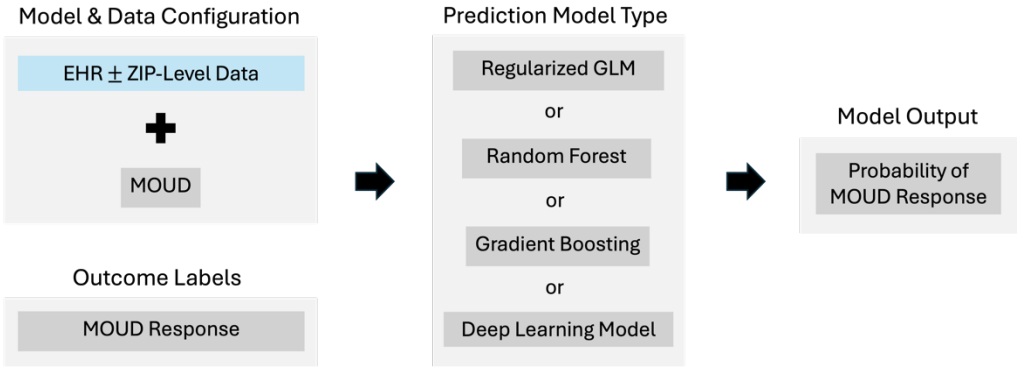
Modeling Plan and Additional Analyses for MOUD Treatment Selection Model

This appendix contains additional information on the methods and analyses for the MOUD treatment response models described in Chapter 4. **Figure 11** shows the different model and data configurations tested. **Figure 12** details the Matthew's Correlation Coefficient. **Figure 13** and **Figure 14** are plots for model discrimination metrics, and **Table 8** contains calibration metrics for EHR-based models. **Figure 15** compares the distributions of differential MOUD treatment response, and **Figure 16** shows MOUD response predictions for an individual who was initiated on buprenorphine.

Figure 11: Model and data configurations for the different MOUD treatment selection models tested.

We adapted the modeling framework from Sheu et al. to also consider different combinations of EHR and ZIP-level data, as shown in blue in (A). The specifics of the early and late fusion models are shown in (B).

(A)



GLM: generalized linear model.

(B)

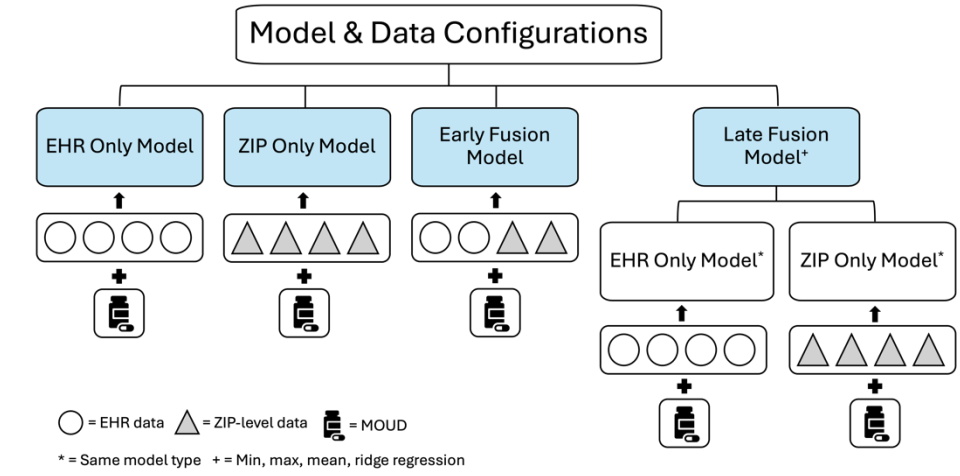


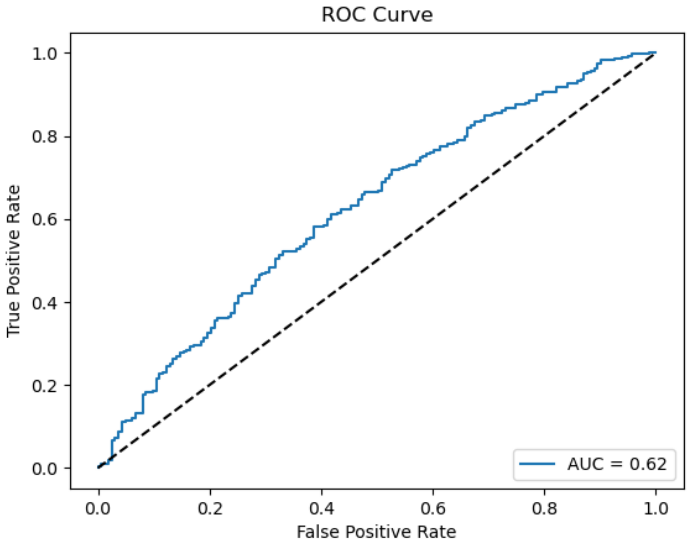
Figure 12: Matthew's Correlation Coefficient formula.

The Matthew's Correlation Coefficient (MCC) measures the correlation of true classes with predicted labels[121]. Values range from -1 (worst) to 1 (best).

$$\text{Matthew's Correlation Coefficient (MCC)} = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

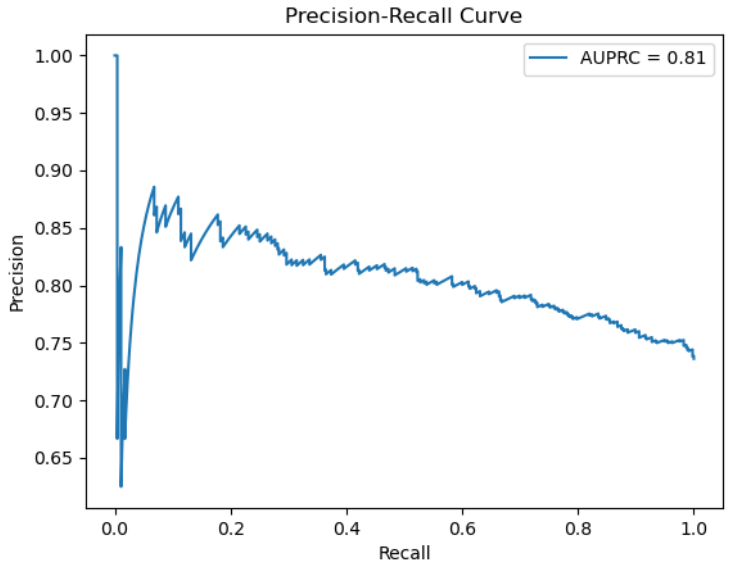
where TP = true positive, TN = true negative, FP = false positive, FN = false negative

Figure 13: Receiver operating characteristic curve for the deep learning model with EHR data only.



ROC: receiver operating characteristic; AUC: area under the curve.

Figure 14: Precision-recall curve for the deep learning model with EHR data only.



AUPRC: area under the precision-recall curve.

Table 8: Calibration metrics for MOUD treatment selection models using EHR data only.

Model	Brier	Intercept	Slope	Spiegelhalter z-score	Spiegelhalter p-value
Random Forest	0.187 [0.166,0.204]	0.276 [-0.0895,0.589]	0.777 [0.513,1.03]	0.649 [-1.66,2.89]	0.356 [0.004,0.961]
Gradient-Boosted Machine	0.186 [0.171,0.2]	0.188 [-0.0846,0.577]	0.828 [0.54,1.1]	0.629 [-1.26,2.32]	0.428 [0.0202,0.976]
Deep Learning Model	0.189 [0.175,0.205]	-1.58 [-3.82,-0.344]	2.67 [1.56,4.62]	-0.942 [-3.65,2.17]	0.194 [0.000264,0.934]

Figure 15: Differential MOUD response probabilities for individuals initiated on each MOUD, all clinicians in test set.

Medians were not significantly different by the Wilcoxon rank-sum test ($p > 0.05$).

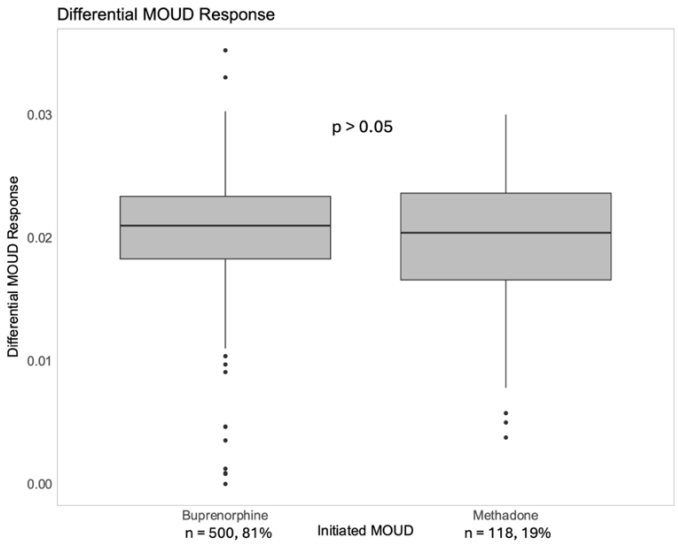
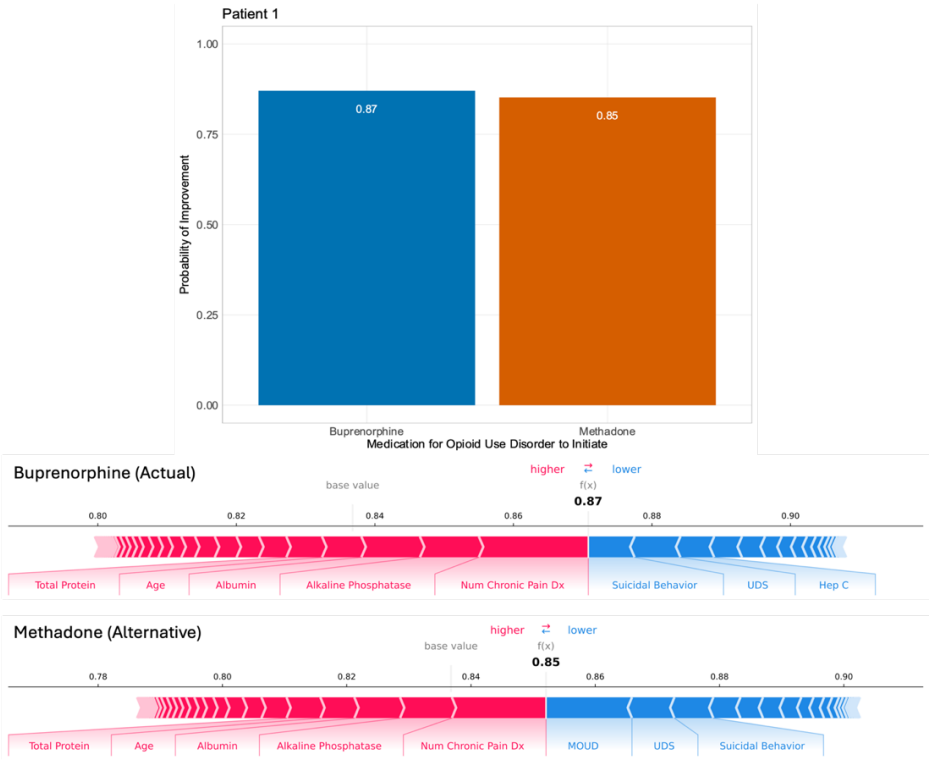


Figure 16: MOUD response probabilities and top predictors for an individual who was prescribed buprenorphine.



Num Chronic Pain Dx: number of chronic pain conditions; Hep C: Hepatitis C; UDS: urine drug screen.

APPENDIX C

Conceptualization and Expansion of Sittig and Singh Socio-Technical Model

This appendix contains more details about the methods and results for the clinician interviews described in Chapter 5. **Figure 17** represents our conceptualization of the Sittig and Singh socio-technical framework used for analysis of interview data. **Table 8** describes how we expanded the definitions of specific dimensions within the Sittig and Singh model.

Figure 17: Overlapping dimensions of the Sittig and Singh socio-technical model.

This figure represents our conceptualization of how the 8 dimensions of the Sittig and Singh socio-technical model are interconnected.

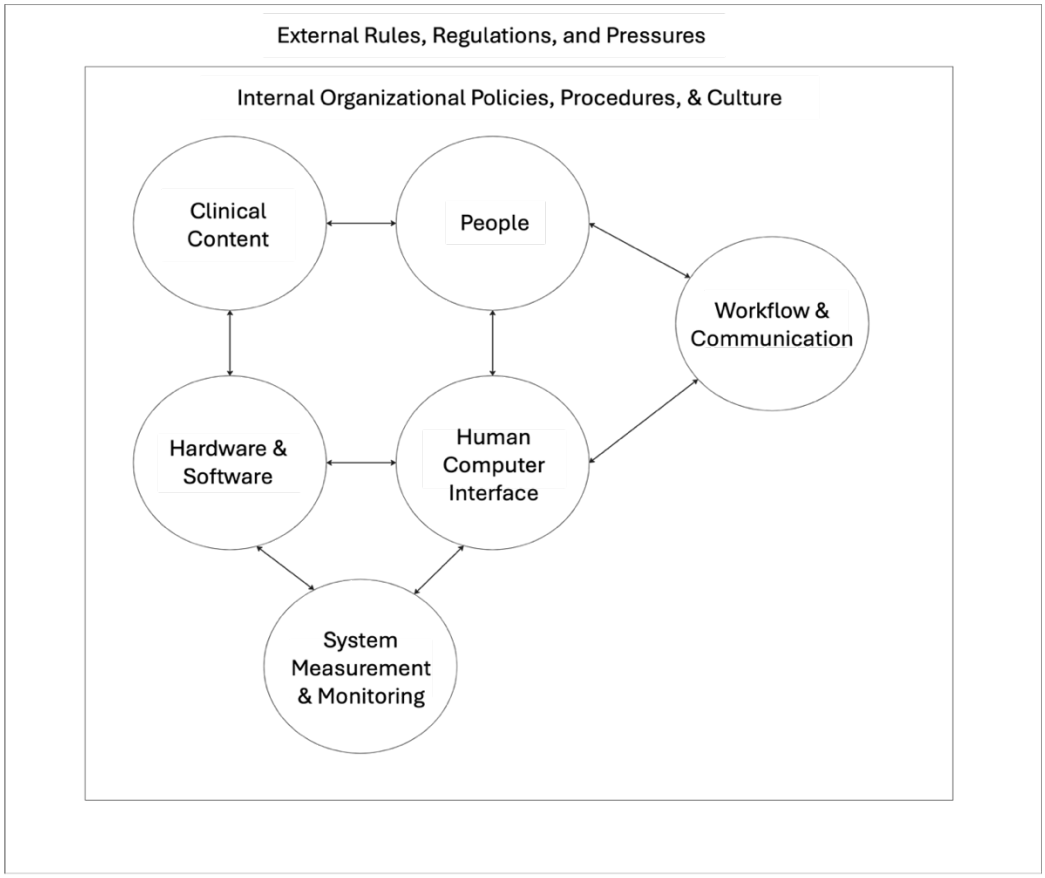


Table 9: Expanded definitions for 3 dimensions in the Sittig and Singh socio-technical model.

Expanded definitions are shown in green.

Socio-Technical Dimension	Expanded Definition
Hardware and Software Computing Infrastructure	Technical aspects of the system. Includes aspects related to development/validation of models.
Clinical Content	Information stored in the system (from clinical standards to patient data). Includes information from internet.
Human Computer Interface	Helps the user engage with the system and interpret information.
People	All individuals involved in design, development, implementation, or use of system.
Workflow & Communication	Collaboration among people in healthcare system. Includes cognitive processes occurring within one person in clinical decision-making.
Internal Organizational Policies, Procedures, & Culture	Institutional rules and structures often developed in response to external factors.
External Rules, Regulations, & Pressures	Factors outside institution that impact design, development, implementation, use, or evaluation of HIT.
System Measurement & Monitoring	Assessing the effects of HIT (e.g., availability of features, use of features, impact on outcomes, unintended consequences).

REFERENCES

1. Spencer M, Garnett M, Miniño A. Drug Overdose Deaths in the United States, 2002-2022. National Center for Health Statistics (U.S.); 2023 Dec.

2. Winhusen T, Walley A, Fanucchi LC, et al. The Opioid-overdose Reduction Continuum of Care Approach (ORCCA): Evidence-based practices in the HEALing Communities Study. Drug and Alcohol Dependence. 2020 Dec 1;217:108325.

3. National Center for Health Statistics. Drug Overdose Mortality by State. CDC/National Center for Health Statistics.

4. Rosenbaum JE, Stillo M, Graves N, et al. Timeliness of provisional United States mortality data releases during the COVID-19 pandemic: delays associated with electronic death registration system and weekly mortality. J Public Health Pol. 2021 Dec;42(4):536–49.

5. Ferris LM, Saloner B, Krawczyk N, et al. Predicting Opioid Overdose Deaths Using Prescription Drug Monitoring Program Data. American Journal of Preventive Medicine. 2019 Dec 1;57(6):e211–7.

6. Saloner B, Chang HY, Krawczyk N, et al. Predictive Modeling of Opioid Overdose Using Linked Statewide Medical and Criminal Justice Data. JAMA Psychiatry. 2020 Nov 1;77(11):1155–62.

7. Gellad WF, Yang Q, Adamson KM, et al. Development and validation of an overdose risk prediction tool using prescription drug monitoring program data. Drug and Alcohol Dependence. 2023 May 1;246:109856.

8. Geissert P, Hallvik S, Van Otterloo J, et al. High-risk prescribing and opioid overdose: prospects for prescription drug monitoring program–based proactive alerts. PAIN. 2018 Jan;159(1):150–6.

9. Ripperger M, Lotspeich SC, Wilimitis D, et al. Ensemble learning to predict opioid-related overdose using statewide prescription drug monitoring program and hospital discharge data in the state of Tennessee. Journal of the American Medical Informatics Association. 2022 Jan 1;29(1):22–32.

10. Dowell D, Brown S, Gyawali S, et al. Treatment for Opioid Use Disorder: Population Estimates — United States, 2022. MMWR Morb Mortal Wkly Rep. 2024 Jun 27;73(25):567–74.

11. Timko C, Schultz NR, Cucciare MA, et al. Retention in medication-assisted treatment for opiate dependence: A systematic review. J Addict Dis. 2016;35(1):22–35.

12. Kast KA, Le TDV, Stewart LS, et al. Impact of inpatient addiction psychiatry consultation on opioid use disorder outcomes. American J Addict. 2024 Mar 28;ajad.13540.

13. Jakubowski A, Singh-Tan S, Torres-Lockhart K, et al. Addiction Consult Service and Inpatient Outcomes Among Patients with OUD. J GEN INTERN MED. 2024 Aug 13;

14. Hasan MM, Young GJ, Shi J, et al. A machine learning based two-stage clinical decision support system for predicting patients' discontinuation from opioid use disorder treatment: retrospective observational study. BMC Med Inform Decis Mak. 2021 Nov 26;21(1):331.

15. Stafford C, Marrero WJ, Naumann RB, et al. Identifying key risk factors for premature discontinuation of opioid use disorder treatment in the United States: A predictive modeling study. Drug and Alcohol Dependence. 2022 Aug 1;237:109507.

16. Gottlieb A, Yatsco A, Bakos-Block C, et al. Machine Learning for Predicting Risk of Early Dropout in a Recovery Program for Opioid Use Disorder. Healthcare (Basel). 2022 Jan 25;10(2):223.

17. Roberts A. Notes from the Field: Nitazene-Related Deaths — Tennessee, 2019–2021. MMWR Morb Mortal Wkly Rep. 2022;71.

18. Korona-Bailey J, Onyango E, Hall KF, et al. Xylazine-Involved Fatal and Nonfatal Drug Overdoses in Tennessee From 2019 to 2022. JAMA Netw Open. 2023 Jul 18;6(7):e2324001.

19. Hershey HL, Onyango EM, Durst K, et al. Tianeptine-involved emergency department visits, fatal overdoses, and substance seizures in Tennessee, 2021–2023. Drug and Alcohol Dependence Reports. 2024 Sep;12:100272.

20. Tennessee Department of Health Office of Informatics and Analytics. 2022 Tennessee Drug Overdose Deaths. Tennessee Department of Health; 2024.

21. Lasko TA, Strobl EV, Stead WW. Why do probabilistic clinical models fail to transport between sites. npj Digit Med. 2024 Mar 1;7(1):1–8.

22. Sittig DF, Singh H. A New Socio-technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems. Qual Saf Health Care. 2010 Oct;19(Suppl 3):i68–74.

23. National Institute on Drug Abuse, National Institutes of Health, U.S. Department of Health and Human Services. Prescription Opioids DrugFacts | National Institute on Drug Abuse (NIDA). 2021.

24. Diagnostic and statistical manual of mental disorders. Vol. 21. Washington, DC: American Psychiatric Association; 2013.

25. Max MB. Quality Improvement Guidelines for the Treatment of Acute Pain and Cancer Pain. JAMA. 1995 Dec 20;274(23):1874.

26. Veterans Health Administration. Pain as the 5th Vital Sign Toolkit. Department of Veterans Affairs; 2000.

27. Phillips DM. JCAHO Pain Management Standards Are Unveiled. JAMA. 2000 Jul 26;284(4):428.

28. Porter J, Hershel J. Addiction Rare in Patients Treated with Narcotics. N Engl J Med. 1980 Jan 10;302(2):123–123.

29. Reuters. OxyContin maker Purdue Pharma stops promoting opioids, cuts sales staff. The Washington Post. 2018 Feb 10;

30. Office of the Attorney General. AG Ferguson: Opioid distributors to pay $518 million to Washington. Washington State. 2022.

31. Office of the Attorney General. AG Ferguson: Johnson & Johnson to pay nearly $150M over its role fueling opioid epidemic. Washington State Office of the Attorney General. 2024.

32. Paulozzi LJ, Budnitz DS, Xi Y. Increasing deaths from opioid analgesics in the United States. Pharmacoepidemiology and Drug. 2006 Sep;15(9):618–27.

33. Rudd RA, Paulozzi LJ, Bauer MJ, et al. Increases in heroin overdose deaths - 28 States, 2010 to 2012. MMWR Morb Mortal Wkly Rep. 2014 Oct 3;63(39):849–54.

34. Gladden RM. Fentanyl Law Enforcement Submissions and Increases in Synthetic Opioid–Involved Overdose Deaths — 27 States, 2013–2014. MMWR Morb Mortal Wkly Rep. 2016;65.

35. United States Department of Health and Human Services. HHS Acting Secretary Declares Public Health Emergency to Address National Opioid Crisis. United States Department of Health and Human Services; 2017.

36. Sprague Martinez L, Rapkin BD, Young A, et al. Community engagement to implement evidence-based practices in the HEALing communities study. Drug and Alcohol Dependence. 2020 Dec;217:108326.

37. Substance Abuse and Mental Health Services Administration. Key Substance Use and Mental Health Indicators in the United States: Results from the 2023 National Survey on Drug Use and Health. United States Department of Health and Human Services; 2024.

38. Dowell D, Ragan KR, Jones CM, et al. CDC Clinical Practice Guideline for Prescribing Opioids for Pain — United States, 2022. MMWR Recomm Rep. 2022 Nov 4;71(3):1–95.

39. Zhang Y, Johnson P, Jeng PJ, et al. First Opioid Prescription and Subsequent High-Risk Opioid Use: a National Study of Privately Insured and Medicare Advantage Adults. J GEN INTERN MED. 2018 Dec;33(12):2156–62.

40. Linas BP, Savinkina A, Madushani RWMA, et al. Projected Estimates of Opioid Mortality After Community-Level Interventions. JAMA Network Open. 2021 Feb 15;4(2):e2037259.

41. Bohnert ASB. Association Between Opioid Prescribing Patterns and Opioid Overdose-Related Deaths. JAMA. 2011 Apr 6;305(13):1315.

42. Jeffery MM, Hooten WM, Hess EP, et al. Opioid Prescribing for Opioid-Naive Patients in Emergency Departments and Other Settings: Characteristics of Prescriptions and Association With Long-Term Use. Annals of Emergency Medicine. 2018 Mar;71(3):326-336.e19.

43. Zedler BK, Saunders WB, Joyce AR, et al. Validation of a Screening Risk Index for Serious Prescription Opioid-Induced Respiratory Depression or Overdose in a US Commercial Health Plan Claims Database. Pain Medicine. 2018 Jan 1;19(1):68–78.

44. Drug Overdose Surveillance and Epidemiology. Data from the Drug Overdose Surveillance and Epidemiology (DOSE) System are Used for Action. Centers for Disease Control and Prevention; 2022.

45. State Unintentional Drug Overdose Reporting System. State Unintentional Drug Overdose Reporting System (SUDORS). Centers for Disease Control and Prevention; 2024.

46. The Network for Public Health Law. State Non-Fatal Overdose Reporting Requirements Fact Sheet. The Network for Public Health Law; 2023.

47. D'Souza RS, Lang M, Eldrige JS. Prescription Drug Monitoring Program. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024.

48. Bamboo Health. NarxCare. Bamboo Health.

49. Gondré-Lewis MC, Abijo T, Gondré-Lewis TA. The Opioid Epidemic: a Crisis Disproportionately Impacting Black Americans and Urban Communities. J Racial and Ethnic Health Disparities. 2023 Aug;10(4):2039–53.

50. Kariisa M, Davis NL, Kumar S, et al. Vital Signs: Drug Overdose Deaths, by Selected Sociodemographic and Social Determinants of Health Characteristics — 25 States and the District of Columbia, 2019–2020. MMWR Morb Mortal Wkly Rep. 2022 Jul 22;71(29):940–7.

51. Hébert AH, Hill AL. Impact of opioid overdoses on US life expectancy and years of life lost, by demographic group and stimulant co-involvement: a mortality data analysis from 2019 to 2022. The Lancet Regional Health - Americas. 2024 Aug;36:100813.

52. Sees KL, Delucchi KL, Masson C, et al. Methadone Maintenance vs 180-Day Psychosocially Enriched Detoxification for Treatment of Opioid Dependence: A Randomized Controlled Trial. JAMA. 2000 Mar 8;283(10):1303.

53. Wakeman SE, Larochelle MR, Ameli O, et al. Comparative Effectiveness of Different Treatment Pathways for Opioid Use Disorder. JAMA Network Open. 2020 Feb 5;3(2):e1920622.

54. Mattick RP, Breen C, Kimber J, et al. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Cochrane Database of Systematic Reviews. 2014;(2).

55. Santo T, Clark B, Hickman M, et al. Association of Opioid Agonist Treatment With All-Cause Mortality and Specific Causes of Death Among People With Opioid Dependence: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2021 Sep 1;78(9):979.

56. Singh D, Saadabadi A. Naltrexone. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024.

57. Sullivan MA, Garawi F, Bisaga A, et al. Management of relapse in naltrexone maintenance for heroin dependence. Drug and Alcohol Dependence. 2007 Dec;91(2–3):289–92.

58. Degenhardt L, Larney S, Kimber J, et al. Excess mortality among opioid-using patients treated with oral naltrexone in A ustralia. Drug and Alcohol Review. 2015 Jan;34(1):90–6.

59. Schuckit MA. Treatment of Opioid-Use Disorders. Longo DL, editor. N Engl J Med. 2016 Jul 28;375(4):357–68.

60. Ma J, Bao YP, Wang RJ, et al. Effects of medication-assisted treatment on mortality among opioids users: a systematic review and meta-analysis. Mol Psychiatry. 2019 Dec;24(12):1868–83.

61. The ASAM National Practice Guideline for the Treatment of Opioid Use Disorder: 2020 Focused Update. J Addict Med. 2020;14(2S Suppl 1):1–91.

62. Dahan A, Yassen A, Romberg R, et al. Buprenorphine induces ceiling in respiratory depression but not in analgesia. British Journal of Anaesthesia. 2006 May;96(5):627–32.

63. Kreek MJ, Vocci FJ. History and current status of opioid maintenance treatments: blending conference session. J Subst Use Addict Treat. 2002 Sep 1;23(2):93–105.

64. Durrani M, Bansal K. Methadone. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024.

65. Huskamp HA, Riedel LE, Barry CL, et al. Coverage of Medications that Treat Opioid Use Disorder and Opioids for Pain Management in Marketplace Plans, 2017. Med Care. 2018 Jun;56(6):505–9.

66. Shulman M, Wai JM, Nunes EV. Buprenorphine Treatment for Opioid Use Disorder: An Overview. CNS Drugs. 2019 Jun;33(6):567–80.

67. 117th Congress (2021-2022). S.445 - Mainstreaming Addiction Treatment Act of 2021. Congress.gov. 2021.

68. 106th Congress (1999-2000): H.R.2634 - Drug Addiction Treatment Act of 2000. 2000.

69. Federal Register. Medications for the Treatment of Opioid Use Disorder. Federal Register; 2024.

70. Jakubowski A, Singh-Tan S, Torres-Lockhart K, et al. Hospital-based clinicians lack knowledge and comfort in initiating medications for opioid use disorder: opportunities for training innovation. Addiction Science & Clinical Practice. 2023 May 18;18(1):31.

71. Degenhardt L, Clark B, Macpherson G, et al. Buprenorphine versus methadone for the treatment of opioid dependence: a systematic review and meta-analysis of randomised and observational studies. Lancet Psychiatry. 2023 Jun;10(6):386–402.

72. Komasi S, Saeidi M, Amiri MM, et al. Triggers of Substance Abuse Slip and Relapse During Outpatient Treatment in Methadone/Buprenorphine Maintenance Therapy Clinics: A Predictive Model with Emphasis on Treatment-Related Factors. Jundishapur J Health Sci. 2017;9(3).

73. Weinstein ZM, Kim HW, Cheng DM, et al. Long-term retention in Office Based Opioid Treatment with buprenorphine. Journal of Substance Abuse Treatment. 2017 Mar;74:65–70.

74. Samples H, Williams AR, Olfson M, et al. Risk factors for discontinuation of buprenorphine treatment for opioid use disorders in a multi-state sample of Medicaid enrollees. Journal of Substance Abuse Treatment. 2018 Dec;95:9–17.

75. Ker S, Hsu J, Balani A, et al. Factors That Affect Patient Attrition in Buprenorphine Treatment for Opioid Use Disorder: A Retrospective Real-World Study Using Electronic Health Records. NDT. 2021 Oct 28;17:3229–44.

76. Biondi BE, Vander Wyk B, Schlossberg EF, et al. Factors associated with retention on medications for opioid use disorder among a cohort of adults seeking treatment in the community. Addiction Science & Clinical Practice. 2022 Mar 7;17(1):15.

77. Moses TE, Rhodes GL, Tavakoli E, et al. Predictors of Retention and Drug Use Among Patients With Opioid Use Disorder Transferred to a Specialty “Second Chance” Methadone Program. Subst Abuse. 2022 Nov 15;16:11782218221138335.

78. Bains N, Abdijadid S. Major Depressive Disorder. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2024.

79. Markou A. Neurobiological Similarities in Depression and Drug Dependence: A Self-Medication Hypothesis. Neuropsychopharmacology. 1998 Mar;18(3):135–74.

80. Rush AJ, Thase ME. Improving Depression Outcome by Patient-Centered Medical Management. AJP. 2018 Dec 1;175(12):1187–98.

81. Lepping P, Whittington R, Sambhi RS, et al. Clinical relevance of findings in trials of CBT for depression. Eur psychiatr. 2017 Sep;45:207–11.

82. Slade EP, Jahn DR, Regenold WT, et al. Association of Electroconvulsive Therapy With Psychiatric Readmissions in US Hospitals. JAMA Psychiatry. 2017 Aug 1;74(8):798.

83. Practice Guideline for the Treatment of Patients With Major Depressive Disorder. American Psychiatric Association; 2010. 152 p.

84. Arnold PIM, Janzing JGE, Hommersom A. Machine learning for antidepressant treatment selection in depression. Drug Discovery Today. 2024 Aug;29(8):104068.

85. Sheu Y han, Magdamo C, Miller M, et al. AI-assisted prediction of differential response to antidepressant classes using electronic health records. npj Digit Med. 2023 Apr 26;6(1):1–12.

86. Berner ES, La Lande TJ. Overview of Clinical Decision Support Systems. In: Berner ES, editor. Clinical Decision Support Systems. New York, NY: Springer New York; 2007. p. 3–22. (Hannah KJ, Ball MJ, editors. Health Informatics).

87. Kouri A, Yamada J, Lam Shin Cheung J, et al. Do providers use computerized clinical decision support systems? A systematic review and meta-regression of clinical decision support uptake. Implement Sci. 2022 Mar 10;17:21.

88. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci. 2017 Sep 15;12(1):113.

89. Osheroff JA 1959, Pifer EA, Teich JM, et al. Improving outcomes with clinical decision support : an implementer’s guide. Chicago, IL: HIMSS; 2005.

90. Davis SE, Embí PJ, Matheny ME. Sustainable deployment of clinical prediction tools—a 360° approach to model maintenance. Journal of the American Medical Informatics Association. 2024 Apr 19;31(5):1195–8.

91. Osheroff JA, Teich JM, Middleton B, et al. A Roadmap for National Action on Clinical Decision Support. Journal of the American Medical Informatics Association. 2007 Mar 1;14(2):141–5.

92. Benda NC, Das LT, Abramson EL, et al. “How did you get to this number?” Stakeholder needs for implementing predictive analytics: a pre-implementation qualitative study. J Am Med Inform Assoc. 2020 Mar 6;27(5):709–16.

93. Weiner SG, El Ibrahimi S, Hendricks MA, et al. Factors Associated With Opioid Overdose After an Initial Opioid Prescription. JAMA Netw Open. 2022 Jan 28;5(1):e2145691.

94. Korona-Bailey JA, Nechuta S, Golladay M, et al. Characteristics of fatal opioid overdoses with stimulant involvement in Tennessee: A descriptive study using 2018 State Unintentional Drug Overdose Reporting System Data. Annals of Epidemiology. 2021 Jun 1;58:149–55.

95. Manasco AT, Griggs C, Leeds R, et al. Characteristics of state prescription drug monitoring programs: a state-by-state survey. Pharmacoepidemiology and Drug. 2016 Jul;25(7):847–51.

96. McElfresh DC, Chen L, Oliva E, et al. A call for better validation of opioid overdose risk algorithms. Journal of the American Medical Informatics Association. 2023 Sep 25;30(10):1741–6.

97. Oliva J. Dosing Discrimination: Regulating PDMP Risk Scores. SSRN Journal. 2021;

98. Valdez A, Cepeda A, Frankeberger J, et al. The opioid epidemic among the Latino population in California. Drug and Alcohol Dependence Reports. 2022 Mar;2:100029.

99. Urban Indian Health Institute. Data Genocide of American Indians and Alaska Natives in COVID-19 Data. Urban Indian Health Institute.

100. Nechuta S, Mukhopadhyay S, Krishnaswami S, et al. Record Linkage Approaches Using Prescription Drug Monitoring Program and Mortality Data for Public Health Analyses and Epidemiologic Studies. Epidemiology. 2020 Jan;31(1):22–31.

101. Tennessee Department of Health Office of Informatics and Analytics. 2022 Drug Overdose Hospital Discharges in Tennessee. Tennessee Department of Health; 2024.

102. Tennessee Department of Health Office of Informatics and Analytics. Tennessee State Unintentional Drug Overdose Reporting System (SUDORS) Report 2024. Tennessee Department of Health; 2024.

103. McPheeters ML, Nechuta S, Miller S, et al. Prescription Drug Overdose Report, 2018. TN Department of Health; 2018.

104. Office of Informatics and Analytics. Data Dashboard. TN Department of Health.

105. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Medicine. 2019 Dec 16;17(1):230.

106. Nelson SD, McCoy AB, Rector H, et al. Assessment of a Naloxone Coprescribing Alert for Patients at Risk of Opioid Overdose: A Quality Improvement Project. Anesthesia & Analgesia. 2022 Mar 28;

107. Dowell D, Ragan KR, Jones CM, et al. CDC Clinical Practice Guideline for Prescribing Opioids for Pain — United States, 2022. MMWR Recomm Rep. 2022 Nov 4;71(3):1–95.

108. Diagnostic and Statistical Manual of Mental Disorders. Vol. 21. Washington, DC: American Psychiatric Association; 541 p.

109. Dowell D, Brown S, Gyawali S, et al. Treatment for Opioid Use Disorder: Population Estimates — United States, 2022. MMWR Morb Mortal Wkly Rep. 2024 Jun 27;73(25):567–74.

110. Tang LA, Gomez M, Suresh U, et al. Clinician Perspectives on a Predictive Model for Recommending Opioid Use Disorder Treatment. AMIA Annu Symp Proc 2024. Forthcoming;

111. Vandenberk B, Vandael E, Robyns T, et al. Which QT Correction Formulae to Use for QT Monitoring? JAHA. 2016 Jun 13;5(6):e003264.

112. Child CG, Turcotte JG. Surgery and portal hypertension. Major Probl Clin Surg. 1964;1:1–85.

113. Shanahan CW, Beers D, Alford DP, et al. A Transitional Opioid Program to Engage Hospitalized Drug Users. J GEN INTERN MED. 2010 Aug;25(8):803–8.

114. Walsh C, Hripcsak G. The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions. Journal of Biomedical Informatics. 2014 Dec 1;52:418–26.

115. Kind AJH, Buckingham WR. Making Neighborhood-Disadvantage Metrics Accessible — The Neighborhood Atlas. N Engl J Med. 2018 Jun 28;378(26):2456–8.

116. United States Census Bureau. Economic Census Tables: 2017. Economic Census Tables;

117. Rickles M, Rebeiro PF, Sizemore L, et al. Tennessee's In-state Vulnerability Assessment for a "Rapid Dissemination of Human Immunodeficiency Virus or Hepatitis C Virus Infection" Event Utilizing Data About the Opioid Epidemic. Clinical Infectious Diseases. 2018 May 17;66(11):1722–32.

118. United States Census Bureau. 2016 ACS 5-Year Data Profile. 2016 Data Profiles American Community Survey;

119. United States Department of Agriculture: Economic Research Service. Rural-Urban Continuum Codes. USDA ERS—Rural-Urban Continuum Codes; 2020.

120. Flanagan BE, Gregory EW, Hallisey EJ, et al. A Social Vulnerability Index for Disaster Management. Journal of Homeland Security and Emergency Management. 2011 Jan 5;8(1).

121. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Mining. 2023 Feb 17;16(1):4.

122. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng. 2018 Oct 10;2(10):749–60.

123. Archambault L, Bertrand K, Martel MO, et al. The current state of knowledge on care for co-occurring chronic pain and opioid use disorder: A scoping review. Journal of Clinical Nursing. 2024 Aug;33(8):3056–76.

124. Soyka M, Zingg C, Koller G, et al. Retention rate and substance use in methadone and buprenorphine maintenance therapy and predictors of outcome: results from a randomized study. International Journal of Neuropsychopharmacology. 2008 Aug 1;11(5):641–53.

125. Damian AJ, Mendelson T, Agus D. Predictors of buprenorphine treatment success of opioid dependence in two Baltimore City grassroots recovery programs. Addict Behav. 2017 Oct;73:129–32.

126. Hser YI, Saxon AJ, Huang D, et al. Treatment retention among patients randomized to buprenorphine/naloxone compared to methadone in a multi-site trial. Addiction. 2014 Jan 1;109(1):79–87.

127. Krawczyk N, Williams AR, Saloner B, et al. Who stays in medication treatment for opioid use disorder? A national study of outpatient specialty treatment settings. Journal of Substance Abuse Treatment. 2021 Jul 1;126:108329.

128. Reuter Q, Larkin GL, Dubé M, et al. Stage-of-change Assessment Predicts Short-term Treatment Engagement for Opioid Use Disorder Patients Initiated on Buprenorphine. West J Emerg Med. 2022 Sep;23(5):684–92.

129. Opioid Treatment Program Directory. Substance Abuse and Mental Health Services Administration.

130. Domzaridou E, Carr MJ, Webb RT, et al. Non-fatal overdose risk during and after opioid agonist treatment: A primary care cohort study with linked hospitalisation and mortality records. Lancet Reg Health Eur. 2022 Nov;22:100489.

131. Substance Abuse and Mental Health Services Administration. Waiver Elimination for Medication-Assisted Treatment (MAT) Act. SAMHSA. 2023.

132. Velez CM, Nicolaidis C, Korthuis PT, et al. “It’s been an Experience, a Life Learning Experience”: A Qualitative Study of Hospitalized Patients with Substance Use Disorders. J Gen Intern Med. 2017 Mar 1;32(3):296–303.

133. Marcovitz DE, Pettapiece-Phillips M, Kast KA, et al. Implementation of a Hub-and-Spoke Partnership for Opioid Use Disorder Treatment in a Medicaid Nonexpansion State. Psychiatr Serv. 2022 Jul;73(7):819–22.

134. Centers for Medicare & Medicaid Services. Maternal Opioid Misuse (MOM) Model | CMS. CMS.

135. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Soc Sci Med. 2022 Jan;292:114523.

136. Wells PS, Anderson DR, Rodger M, et al. Evaluation of D-Dimer in the Diagnosis of Suspected Deep-Vein Thrombosis. N Engl J Med. 2003 Sep 25;349(13):1227–35.

137. Miech EJ, Rattray NA, Flanagan ME, et al. Inside help: An integrative review of champions in healthcare-related implementation. SAGE Open Med. 2018 May 17;6:2050312118773261.

138. Savage T, Nayak A, Gallo R, et al. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. NPJ Digit Med. 2024 Jan 24;7(1):1–7.

139. Varonen H, Kortteisto T, Kaila M, et al. What may help or hinder the implementation of computerized decision support systems (CDSSs): a focus group study with physicians. Fam Pract. 2008 Jun 1;25(3):162–7.

140. Cook LA, Sachs J, Weiskopf NG. The quality of social determinants data in the electronic health record: a systematic review. J Am Med Inform Assoc. 2022 Jan 1;29(1):187–96.

141. Cohen SM, Joab R, Bolles KM, et al. Ending Medical Complicity With Skilled-Nursing Facility Discrimination Against People With Opioid Use Disorder. Ann Intern Med. 2023 Mar 21;176(3):410–2.

142. Benda NC, Novak LL, Reale C, et al. Trust in AI: why we should be designing for APPROPRIATE reliance. J Am Med Inform Assoc. 2022 Jan 1;29(1):207–12.

143. Embi PJ. Algorithmovigilance—Advancing Methods to Analyze and Monitor Artificial Intelligence–Driven Health Care for Effectiveness and Equity. JAMA Netw Open. 2021 Apr 15;4(4):e214622.

144. Kaiser Family Foundation. Status of State Medicaid Expansion Decisions: Interactive Map. KFF. 2024.

145. Lo-Ciganic WH, Donohue JM, Yang Q, et al. Developing and validating a machine-learning algorithm to predict opioid overdose in Medicaid beneficiaries in two US states: a prognostic modelling study. Lancet Digit Health. 2022 Jun;4(6):e455–65.

146. Electronic Medical Records and Genomics (eMERGE) Network.

147. University of California, San Francisco, Adler-Milstein J, Aggarwal N, et al. Meeting the Moment: Addressing Barriers and Facilitating Clinical Adoption of Artificial Intelligence in Medical Diagnosis. NAM Perspectives. 2022 Sep 29;22(9).