

# Understanding the Opioid Epidemic: Human-Based Versus Algorithmic-Based Perceptions, Treatments, and Guidelines

Alireza Boloori

Milgard School of Business, University of Washington, Tacoma, aboloori@uw.edu

Soroush Saghafian

Harvard Kennedy School, Harvard University, Soroush.Saghafian@hks.harvard.edu

Stephen J. Traub

Department of Emergency Medicine, Brown University, stephen.traub@brownphysicians.org

As a major public health crisis, the opioid epidemic caused over 556,000 deaths in the U.S. between 2000 and 2020. To control the epidemic, the Centers for Disease Control and Prevention (CDC) has developed some general guidelines, encouraging physicians to use opioid medications only when their benefits outweigh their risks. The CDC's 2016 guidelines mainly left it to physicians to decide when the benefits outweigh the risks. A few years later (in 2022), the CDC made some modifications to make its recommendations a bit less reliant on each individual physician's perception of benefits versus risks. In complex and high stake decision-making environments such as those pertaining the use of opioid medications, it is not clear whether and how human-based perceptions might differ from algorithmic-based ones. In this study, we first develop some longitudinal machine learning algorithms (e.g., historical random forest, recurrent neural networks, and long short-term memory networks) and train them on clinical evidence of more than 3 million patients. We then feed the best machine learning algorithm to a mathematical model that enables determining cost-effective treatments for each patient in a personalized manner. Through extensive numerical experiments, we compare the treatment options and recommendations from our algorithmic-based approach with human-based ones that are currently followed in the medical practice. Compared to the human-based approach, our results show that the average saving in quality-adjusted life years and costs obtained by following our algorithmic-based treatments are about 2.82 days and \$461.46 per patient per year. Finally, we make use of our findings and generate insights for policymakers as well as individual physicians into better ways of managing opioid prescriptions (and hence, the opioid epidemic) by incorporating and interacting with our algorithmic-based approach.

*Key words:* Human vs. machine; opioid epidemic; pain management; personalized medicine; machine learning

*History:* December 8, 2022

---

## 1. Introduction

According to the National Institute on Drug Abuse (NIDA), 915,515 drug-related deaths occurred in the U.S. between 2000 and 2020, among which opioid analgesics (i.e., painkillers) were the main contributing factor accounting for 556,472 deaths (60.78% of total deaths). These opioid painkillers often result in patients switching to heroin or synthetic opioids (e.g., fentanyl), which, in turn, caused additional 357,423 deaths during the same period (NIDA 2022). The economic cost of the U.S. opioid epidemic was estimated to be \$1,021 billion in 2017 (Luo et al. 2021), with about two million people being either dependent on prescription opioids or abusing them (USA Today 2016).

To address this crisis, the Centers for Disease Control and Prevention (CDC) proposed a set of guidelines in 2016 with the aim of reducing opioid prescriptions by clinicians (CDC 2016, Dowell et al. 2016). The CDC guidelines, however, have been widely criticized for several reasons, including

the fact they did not make use of existing clinical evidence to provide a clear level of specificity. For example, the American Medical Association (AMA) criticized CDC guidelines by stating that “*we continue to have serious concerns that some [of these guidelines] either contain a degree of specificity not supported by the existing evidence or conflict with official Food and Drug Administration (FDA)-approved product labeling for opioid analgesic products.*” (AMA 2016)

In essence, the CDC guidelines heavily relied on human judgment by recommending that “*Clinicians should consider opioid therapy only if expected benefits for both pain and function are anticipated to outweigh risks to the patient*” (Dowell et al. 2016). However, understanding when, and for what patients, the benefits outweigh the risks is not an easy task for human clinicians. Furthermore, the lack of specificity in the CDC recommendations would make them a “one-size-fits-all” approach. For example, instead of providing recommendations tailored to each patient characteristics, the CDC recommended fixed thresholds for all patients: no more than 90 morphine milligram equivalent (MME) per day for the strength of opioids or no more than 7 days of supply for acute pain (CDC 2016).<sup>1</sup> Lack of a clear treatment guideline tailored to each patient characteristics, juxtaposed with *human-based perceptions*, have potentially led clinicians to show wide variations in opioid prescriptions, which, in turn, is another contributing factor to the opioid epidemic (Barnett et al. 2017). A main goal of our study is to use large-scale clinical evidence and develop an analytics-driven *algorithmic-based* approach that can (1) help clinicians (as the human component of the decision-making process) quantify when, and for what patients, the benefits of opioid prescriptions outweighs their risks, and (2) allow policymakers to create recommendations that are personalized (i.e., adapted to each patient’s characteristics) and not based on one-size-fits-all rules.

To this end, we make use of a large-scale claim data—Merative<sup>TM</sup> MarketScan<sup>®</sup> Commercial Dissertation Databases—which contains the history of medical encounters and prescribed medications of over 3 million patients. A summary of our data is shown in Table 1. From this data, we retrieve information on the medical history of each individual patient, including diagnoses made by providers, prescriptions of pharmacologic treatments (opioid and non-opioid medications), and use of non-pharmacologic treatments (e.g., physical therapy and chiropractic). Training longitudinal machine learning algorithms (e.g., historical random forest, recurrent neural networks, and long short-term memory) on this data allows us to predict, based on each individual patient’s characteristics, the risks of (1) opioid dependence, abuse, overdose, or death, and (2) pain remaining untreated or undertreated. We refer to these risks as the risks of *opioid use disorder* (OUD) and *undertreated pain* (UTP), respectively. We then compare these machine learning algorithms, and feed the one that has the most accurate risk predictions to a mathematical model that allows

<sup>1</sup> In early 2022, the CDC delivered the first draft of a new set of guidelines, where the most notable change from the 2016 guidelines is the removal of the thresholds for the opioid dose and duration of supply. Despite this change, many physicians, patients, and organizations have reservations against these guidelines, stating that the modified guidelines still mostly focus on the harms of opioids rather than their benefits in avoiding poorly managed pain (NPR 2022).

**Table 1** Data summary (left: some of variables in the data, right: patient demographic, risk covariates, etc.)

Variable	Data Table						Variable	Average (S.D.)
	A <sup>1</sup>	I <sup>2</sup>	S <sup>3</sup>	O <sup>4</sup>	D <sup>5</sup>	R <sup>6</sup>		
Patient ID	✓	✓	✓	✓	✓	✓	# patients	3,013,637
Age	✓	✓	✓	✓	✓	✓	# observations (visits & prescriptions)	25,340,400
Gender	✓	✓	✓	✓	✓	✓	# pain medication prescriptions	11,302,380
Monthly Enrollment	✓						Demographics	
Admission date		✓	✓				Age (years) pre-supply	40.02 (15.76)
Discharge date		✓	✓				Fraction <sup>†</sup> gender (female)	57.92% (0.03%)
Discharge status		✓					Pain pre-supply	
Service year		✓	✓	✓	✓		Fraction with acute (no chronic) pain	56.18% (0.03%)
Service place			✓	✓			Fraction with chronic (no acute) pain	42.12% (0.03%)
Service date start				✓			Risk: surgery or inpatient admission	
Service date finish				✓			Fraction with surgical procedures	77.16% (0.02%)
Provider type			✓	✓			Fraction with inpatient admissions	22.87% (0.02%)
Diagnosis codes		✓	✓	✓			Risk: behavioral factors	
Procedure codes		✓	✓	✓			Fraction with alcohol consumption	1.31% (0.01%)
Procedure type				✓			Fraction with smoking	3.98% (0.01%)
Drug name						✓	Fraction with substance abuse	0.49% (0.00%)
Strength						✓	Fraction with non-substance abuse	1.11% (0.01%)
Consumption method						✓	Fraction with mental health disorder	55.45% (0.03%)
Schedule type						✓		
Drug ID					✓	✓		
Therapeutic class					✓	✓		
Days supply					✓			
# units dispensed					✓			
# refills					✓			
Payments		✓	✓	✓	✓			

<sup>1</sup>Annual insurance enrollment of patients. <sup>2</sup>Inpatient admissions. <sup>3</sup>Inpatient services. <sup>4</sup>Outpatient services. <sup>5</sup>Outpatient pharmaceutical claims. <sup>6</sup>Red book: drugs general information.

<sup>†</sup>Fraction is out of the number of patients (3,013,637).

determining cost-effective multi-modal pain treatment (i.e., including both pharmacologic and non-pharmacologic) considering both the benefits and risks of different treatment options. Despite the high accuracy of the machine learning methods in risks predictions, their black-box nature, together with the curse of dimensionality in the optimization model (due to considering multiple treatment modes), may reduce the interpretability of our approach, making it less desirable to understand and follow by clinicians and policymakers. Thus, to facilitate the adoption of our algorithmic-based approach, we develop two heuristic solution methods that are (a) highly interpretable, and (b) easily implementable in a medical decision-support system.

Overall, our contributions are two-fold:

- (1) We develop an algorithmic-based approach capable of dynamically finding multi-modal treatments personalized to individual patients based on their characteristics.
- (2) By training our algorithm on large-scale clinical data, we provide important insights and implications for both policymakers and individual physicians:
  - By comparing the cost-effectiveness of our algorithmic-based treatment plans with the CDC 2016 and 2022 guidelines, we find that the modifications the CDC has made in its 2022 guidelines would improve the cost-effectiveness of interventions adopted by physicians. However, we find that the CDC 2022 guidelines are still not as effective as our algorithmic-based approach, and

hence, provide insights into ways the guidelines could be further improved. As the CDC is planning to soon finalize the evaluation of the first draft of its 2022 guidelines (NPR 2022), we hope our insights help the authorities to provide more effective recommendations.

- Our results show that the CDC guidelines should also emphasize the harms of poorly managed pain as much as they target avoiding the harms of using opioids. For example, we find that under the current practice, for each patient experiencing OUD, there are about 24 patients with UTP. The algorithmic-based treatment plans we obtain provide a much better balance between OUD and UTP, and hence, can reduce this number to about 10.
- Our algorithmic-based approach would improve the average quality-adjusted life years (QALYs) gained and cost incurred compared to the human-based approach (in which a human-based judgment is used to assess the potential benefits and risks) by 2.82 days and \$461.46 per patient per year, respectively. We also find that patients with acute pain, or those with a history of behavioral factors (e.g., mental health disorders or substance abuse) or surgeries/inpatient admissions would benefit most from the treatment plans we obtain.
- Our results show that, compared to the current human-based approach, opioids could be prescribed at a higher intensity for patients with no history of behavioral factors. Among this group, the opioid dose can be tapered down faster for younger females (compared to older males) or patients with a record of surgeries/inpatient admissions. Also, to balance the opioid dose being tapered down, the duration of opioid medications should be increased after the onset of pain, and then be decreased towards the end of the therapeutic course. Furthermore, in the presence of behavioral factors or surgeries/inpatient admissions, older males can be prescribed with a higher duration of opioid medications. However, in their absence, increasing the duration of such medications is mainly useful only for younger females.
- Finally, we find that the rate of using non-pharmacologic treatments (NPHTs), such as physical therapy or chiropractic, needs to be increased. In particular, patients with acute pain could use NPHTs for a few months after the onset of pain, while any record of behavioral factors would elongate this period. For patients with chronic pain, however, NPHTs should be used over the whole (only towards the end of) time horizon when there is a (is no) history of such factors.

The rest of this paper is organized as follows. In §2, we provide a review of relevant literature. In §3, we discuss our data and study design. In §4, we present our analytical setting, including the longitudinal machine learning models we have trained to predict the risks of OUD and UTP for each patient, as well as a mathematical model we have developed to obtain cost-effective multi-modal pain treatment plans. In §5, we present our numerical experiments and discuss relevant insights and implications. In §6, we provide a summary of our policy recommendations. Finally, in §7, we conclude the paper and briefly discuss avenues for future research.

## 2. Literature Review

### 2.1. Studies on Side Effects of Opioid Medications

A stream of literature related to our work focuses on establishing statistical associations between opioid-related side effects and various risk factors such as age, gender, duration of supply, high dose of opioid, and history of alcohol abuse, smoking, and mental health disorders (see, e.g., Cochran et al. (2014) and Ciesielski et al. (2016)). For reviews of studies on the impact of opioid painkillers on addiction, drug dependence, overdose, or death, one can refer to Fishbain et al. (2007) and Nuckols et al. (2014).

Within this stream, some studies apply machine learning to address the opioid-related side effects. Haller et al. (2016) employed Natural Language Processing techniques to predict risks of drug abuse and addiction before a prescription is written. Che et al. (2017) made use of a Deep Feed-Forward Neural Network to predict the possibility of long-term opioid use. Crosier et al. (2017) used random forests to predict opioid overdose. Vunikili et al. (2018) utilized an Extreme Gradient Boosting algorithm along with logistic regression to predict the risk of opioid abuse, overdose, and death. Bjarnadottir et al. (2019) applied LASSO, adaptive boosting, and random forests to establish risk factors associated with the chronic use of opioid painkillers. Compared to this stream, we not only analyze the risk of opioid dependence, abuse, overdose, and death, but also measure the risk of undertreated pain to account for potential benefits of pain treatments. Furthermore, the foregoing studies are based on *cross-sectional* analyses (a patient's record is gathered at a single point in time without considering the dynamics of patient behavior or health information), whereas we take a *longitudinal* approach. This allows us to study, not only for *what* patients, but also *when* the benefits of opioid medications overcome their risks.

### 2.2. Studies on Measuring Pain and Efficacy of Pain Management

To measure the intensity of pain and evaluate the efficacy of managing it, there are evidence-based pain assessment scales, such as verbal rating scales, numerical rating scales, visual analogue scales, and behavioral pain scales (see, e.g., Huskisson (1974), Katz and Melzack (1999)). These scales are typically evaluated using patient-based surveys, which are often impacted by patients' satisfaction (see, e.g., Wells et al. (2008)). However, accounting for patients' satisfaction is known to inadvertently propell providers to prescribe opioid medications. Thus, the Centers for Medicare and Medicaid Services has recently proceeded towards removing pain management questions from the Hospital Consumer Assessment of Healthcare Providers and Systems survey (see, e.g., Bolori et al. (2020b)). To characterize the efficacy of pain management in our study, we instead focus on repeated visits due to an unresolved pain related condition, which indicates whether the condition is being treated effectively (see, e.g., McPhillips-Tangum et al. (1998) and Xiao and Barber (2008)). In particular, we make use of information from our claims data about repeated visits of patients before and after each opioid prescription (for more details, see §3.3.2).

### 2.3. Relevant Operations Research/Management Science Studies

Pitt et al. (2018) proposed a dynamic compartmental model based on a combination of pain, opioid use, and addiction status, and found that various resource allocation schemes, (e.g., supplying naloxone as an opioid antagonist, supplying needles for addicted patients, and promoting medication-assisted treatments), could have a positive long-term impact on patients' quality of life. Freeman et al. (2019) conducted an empirical study and showed that getting a second opinion such as visiting another provider rather than a primary care provider for opioid prescription is associated with a lower rate of long-term opioid use. Gan et al. (2019) analyzed the impact of wearable devices on detecting opioid use disorder via urine tests, and proposed a partially observable Markov decision process model to optimize decisions on wearing these devices given various budget and patient adherence considerations. A broader set of issues in optimizing decisions regarding care delivery via wearables and mobile health (mHealth) devices is discussed in studies such as Saghafian and Murphy (2021), and the references therein.

From a methodology standpoint, the Operations Research/Management Science (OR/MS) literature has applied supervised learning in optimization frameworks. For example, Bertsimas et al. (2016) followed this approach to improve cancer chemotherapy regimens. However, this body of literature has primarily focused on single-period decision-making problems, where the machine learning is applied to cross-sectional data. Reinforcement Learning (RL) is another branch of machine learning suited for modelling multi-period decision-making problems. For example, Saghafian (2022) shows how RL can be applied to longitudinal observational data sets in order to obtain optimal dynamic treatment regimes that yield casual improvements, even if the longitudinal observational data is subject to hidden confounders. Other approaches in handling multi-period settings include Markov decision processes and multi-armed bandit models, which have been applied to a variety of problems such as in optimizing warfarin dosing (Bastani and Bayati 2020) and joint immunosuppressive and diabetes medications (Boloori et al. 2020a). Despite their widespread applications, these methods could fall short in addressing the curse of dimensionality, nonstationary rewards, and history-dependency that may arise in problems like the one we study in this paper. In addition, the curse of ambiguity introduced in Saghafian (2018, 2022) often limits the applicability of these methods when applied to observational longitudinal data sets.

We contribute to this stream by developing a multi-period optimization model based on *recurrent neural networks* (RNN) that allow us to analyze our longitudinal data. The RNN framework also enables us to address some of the foregoing challenges such as dependency on history and nonstationarity of rewards. To the best of our knowledge, our study is among the first in OR/MS applying longitudinal deep learning to optimization models. From an application standpoint, to the best of our knowledge, our proposed framework is the first to simultaneously address side effects and potential benefits of multi-modal pain treatments. As we show, this is important, as it can

yield superior insights into how (a) physicians as human experts should prescribe these treatments in practice, and (b) policymakers should adjust the guidelines.

Finally, a stream of literature suggests that “algorithm aversion” might limit the impact that using an algorithmic approach such as ours might have in practice, because human experts might be reluctant to make use of recommendations obtained from algorithms (Dietvorst et al. 2015). Some recent studies, however, show that humans do possess “algorithm appreciation” (Logg et al. 2019), and hence, in critical and complex decision-making settings such as the one we study in this work, they are likely to take into account the advice they receive from a well-designed algorithm. We believe improving the interpretability and removing the black-box nature of some of the algorithms can go a long way in this regard. Thus, we adopt solution methodologies that allow reducing the underlying complexities, and enable obtaining interpretable recommendations.

### 3. Data and Study Design

The data that we utilize in our study comes from the Merative MarketScan Commercial Dissertation Databases for commercial claims and encounters (“CCAE” for brevity) for years 2008-2010. There are two main sources of information that we have retrieved from the CCAE databases: (a) information about patients’ encounters and diagnoses,<sup>2</sup> and (b) information about the history of prescriptions.

#### 3.1. Inclusion Criteria

We applied the following inclusion criteria, after which 3,013,637 patients were left in our data set:

- (1) Full insurance enrollment in each year during 2008-2010. The CCAE data contains information about patients (and their dependents) with private insurance coverage, and changes in such coverage can result in a discontinuation of a patient’s medical records.
- (2) No history of cancer or end-of-life (palliative) status. The CDC guidelines on opioid prescription do not apply to patients who suffer from cancer or are in their palliative stage (Dowell et al. 2016). When evaluating the CCAE data, we use the following factors to identify such cases from the ICD-9 diagnoses codes: cancer, neoplasm, malignant, malignancy, benign, carcinoma, and palliative.
- (3) No history of congenital anomalies or conditions originated in the perinatal period.
- (4) At least one episode of an opioid analgesic prescription.
- (5) No opioid prescription within the first 90 days of 2008 (the beginning of our data). Patients whose records satisfy this condition are called *opioid naïve*, and their prescription might be related to a medical condition occurring prior to the beginning of our data. To be consistent with the medical literature (see, e.g., Johnson et al. (2016)), we select a conservative range of 90 days in our analysis.

<sup>2</sup> To determine the diagnoses, we follow the International Classification of Diseases, Ninth Revision, Clinical Modification (referred to as ICD-9 hereafter).

(6) No history of opioid side effects within the first 90 days of 2008. If an overdose occurs for a patient, we must know which medications s/he had used prior to the overdose that might have had significant impacts on this incidence. Therefore, consistent with criterion (5), we make use of a 90-day window to exclude patients without sufficient information within our data set.

(7) No opioid prescription occurring prior to the first recorded encounter. We exclude patients that do not satisfy this criterion, because our aim is to identify a pain-inducing medical condition due to which an opioid has been prescribed for the first time.

### 3.2. Independent Variables

A summary of our independent variables are presented in Table 2. Below, we provide further information on how we have measured these variables.

**3.2.1. Treatments: Pharmacologic.** We measure the strength and duration of supply for both opioid and non-opioid medications. We do this in three consecutive phases described below.

**Phase 1: Information of original prescribed drugs.** We first identify opioid and non-opioid drugs that have been prescribed to patients. We note that some drugs can have both opioid and non-opioid components. For example, ‘Acetaminophen/Propoxyphene’ with strength 325mg-50mg is a drug, where ‘Acetaminophen’ (‘Propoxyphene’) is a non-opioid (opioid) component with the strength of 325 (50) milligram per tablet. For such drugs, we decompose the generic name and differentiate the strength of opioid from that of non-opioid. Out of 50 opioid and 51 non-opioid unique drugs that are prescribed to patients in our data set, we determine 21 opioid and 34 non-opioid unique components. List of these components are provided in Appendix B.1.

To obtain the strength and duration of supply for the identified medications, we employ the following information from the CCAE data: the strength per drug unit (STRENGTH), the number

**Table 2 Summary of independent variables**

Variable <sup>†</sup>	Average (S.D.) <sup>‡</sup>	Type <sup>††</sup>
Demographics: gender (female)	57.92% (0.03%)	Static
Demographics: age (years) before the first opioid prescription	40.02 (15.76)	Dynamic
Treatments–Pharmacologic: strength, opioid (MME)	37.09 (28.26)	Dynamic
Treatments–Pharmacologic: strength, non-opioid (mg)	352.21 (230.70)	Dynamic
Treatments–Pharmacologic: duration of supply (days)	14.70 (10.40)	Dynamic
Treatments–Pharmacologic: use of non-opioid medication pre-supply	14.81% (0.02%)	Static
Treatments–Nonpharmacologic: use	45.05% (0.03%)	Dynamic
Pathology of pain:	(1.71 (0.01), 14.13 (0.02),	Static
(no pain, chronic secondary, chronic primary, acute)	27.99 (0.03), 56.18 (0.03))%	
Behavioral Risk Factors: history of alcohol consumption	1.31% (0.01%)	Dynamic
Behavioral Risk Factors: history of smoking	3.98% (0.01%)	Dynamic
Behavioral Risk Factors: history of substance abuse	0.49% (0.00%)	Dynamic
Behavioral Risk Factors: history of non-substance abuse	1.11% (0.01%)	Dynamic
Behavioral Risk Factors: history of mental health disorder	55.45% (0.03%)	Dynamic
History of surgeries	77.16% (0.02%)	Dynamic
History of inpatient admissions	22.87% (0.02%)	Dynamic
Number of visits in each window (see Remark 1)	1.89 (2.59)	Dynamic

<sup>†</sup>History: the variable’s value will not change once it occurs. <sup>‡</sup>Rates are reported out of 3,013,637 patients considered in our study. <sup>††</sup>Dynamic (static): the variable can (does not) change over time.



of units dispensed per prescription (QTY), the number of refills (REFILL), and days of supply for each refill (DAYSUPP). We then set:

$$\text{Duration of supply} = (\text{REFILL} + 1) * \text{DAYSUPP}, \quad (1a)$$

$$\text{Original strength (per day of supply)} = \text{STRENGTH} * (\text{QTY} / \text{Duration of supply}), \quad (1b)$$

where Equation (1b) is used twice, once for measuring the strength of opioids and once for that of non-opioids. Furthermore, to test the potential collinearity between the resulted strengths and duration of supply, we conduct the Pearson’s product-moment correlation test with the null hypothesis ( $H_0$ ): there is no correlation. Based on our results ( $p$ -value  $> 0.05$ ), we could not reject the following null hypothesis at the 95% confidence level.

**Phase 2: Transformation of drugs’ strength.** We then transfer the original strength of opioid medications obtained from Equation (1b) to a common measure known as the *Morphine Milligram Equivalent* (MME). For example, 1 milligram (mg) of *Oxycodone* is equal to 1.5 MME, but 1 milligram of *Hydromorphone* is equal to 4 MME (see, e.g., Palliative Drugs (2009) and CDC (2020)). For non-opioid medications, we transfer their strengths to a common unit: milligram (mg). In Table 3, we summarize the information related to pain medications obtained after phases 1 and 2 described above.

**Phase 3: Adjustment based on number of prescriptions.** To set up time intervals in which patients visit providers to assess their condition, we consider a time window of  $W$  days starting from the beginning of the first opioid supply. To be consistent with the recommended regular intervals of opioid therapy reassessment (CDC 2016), we set  $W = 30$  in our study. We then adjust the strengths and duration of supply based on the number of times medications are prescribed in each 30-day time window. Figure 1 illustrates an example of this adjustment. It should be noted that different durations of supply are additive. However, different strengths are not additive. For example, in Figure 1, we have a total of  $10+5=15$  days of supply in the time window, but we cannot measure the total strength as  $30+50=80$  MME. In addition to the foregoing variables, and to thoroughly explore the efficacy of pain medications, we consider the history of using non-opioid medications prior to the first opioid supply.

**Table 3 Summary of information related to pain medications after phases 1 and 2**

Description	Value
Number of pain medication prescriptions	11,302,380
Number of refills	Average (standard deviation) = 0.22 (0.91)
Duration of supply (days)	Average (standard deviation) = 18.05 (26.86)
Sstrength, opioid (MME)	Average (standard deviation) = 38.13 (42.40)
Strength, non-opioid (mg)	Average (standard deviation) = 1,041.51 (489.92)
Method of drug consumption (out of all prescriptions)	Oral (99.42%), Transdermal (0.58%)

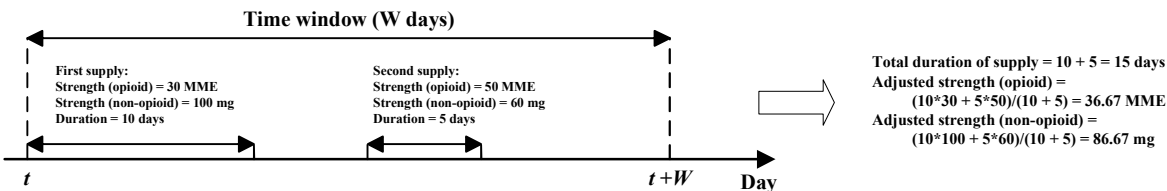


Figure 1 Adjustment of strengths and duration of supply based on the number of prescriptions

**3.2.2. Treatments: Non-Pharmacologic.** A multi-modal pain treatment involves both pharmacologic treatments (e.g., opioid and non-opioid medications) and non-pharmacologic treatments (e.g., physical therapy, chiropractic, and acupuncture). We identify the use of non-pharmacologic treatments from our data set via codes related to procedure groups, provider types, and service categories (see Table 4).

**3.2.3. Pathology of Pain.** We identify the clinical condition(s) based on which an opioid medication is prescribed in the first place. However, in claims data, the information on drugs prescription is not typically linked to visits and corresponding diagnoses. As a result, a pain-inducing condition is not necessarily identifiable. To address this, we focus on the *pre-supply period* defined as the time period prior to the first opioid prescription (i.e., prior to the first time window in our study period).<sup>3</sup> Specifically, we identify acute or chronic pain conditions by exploring ICD-9 diagnoses codes within the pre-supply period. For acute pain, we include acute conditions, injuries, or accidents that are explicitly labeled in ICD-9 codes, or other symptoms that are either indicative of critical conditions, or would typically require immediate medical care or surgical procedures. For brevity, we hereafter refer to these conditions as “acute,” and term the other conditions pertaining pain as “chronic.”

We observe from our data that there are typically two pathological paths for chronic pain that could warrant a pain treatment. In the first path, there is no secondary medical condition involved, and the chronic pain is primarily managed by pain medications (e.g., migraine). In the second path, pain is induced by a secondary condition (e.g., hypertensive heart disease or pneumonia), and addressing this secondary condition (e.g., by prescribing specialty drugs or performing designated treatments) typically precedes prescribing pain medications. We refer to these two paths as *primary* and *secondary* chronic pain conditions, respectively. In Appendix B.2, we represent ICD-9 diagnoses codes that we have used to identify these conditions.

<sup>3</sup> Pain medications are typically supplied within a short period after observing a pain-inducing condition (e.g., surgery). Nevertheless, we consider a pre-supply period of 90 days to be conservative.

Table 4 Codes used to identify non-pharmacologic treatments

<b>Procedure groups:</b> 181–189 (physical medicine), 191 and 195 (chiropractic/spinal manipulation)
<b>Provider types:</b> 120 (chiropractor), 140 (pain management), 350 (physical medicine and rehab), 865 (acupuncturist), 870 (spiritual healers)
<b>Service categories:</b> 3xy18 for $x \in \{0, 1\}$ and $y \in \{1, 2, 3, 4, 5, 6\}$ (behavioral health therapy)

**3.2.4. Behavioral Risk Factors.** We consider five types of behavioral factors, including alcohol consumption, smoking, substance drug disorders, non-substance drug disorders, and mental health disorders. In Appendix B.3, we provide more details on how we have identified these factors using our data.

**3.2.5. Surgeries and Inpatient Admissions.** Finally, we involve information related to whether a patient has had a surgical procedure and/or received an inpatient admission. We do so because the treatment of post-surgical pain with opioid medications has been reported as a cause for an onset of post-operative opioid addiction or long-term opioid usage after hospital discharge (see, e.g., Pletcher et al. (2008) and Wu et al. (2019)). In Appendix B.4, we provide further information on how we have identified surgery and inpatient admissions using our data.

### 3.3. Dependent Variables

Our dependent variables capture the occurrence of two events: opioid use disorder (OUD) and undertreated pain (UTP). We term these as Events 1 and 2, respectively, and define them below. A suitable treatment is one that avoids both of these events.

**3.3.1. Event 1 (OUD).** We define Event 1 so that captures whether the patient experiences any side effect that can be attributed to OUD.

**DEFINITION 1 (EVENT 1).** We say Event 1 has occurred within a time window if, in that window, there is at least one incidence of dependence, abuse, poisoning, or any adverse effect caused by an opioid.

In Appendix C.1, we present the ICD-9 codes that we have used to identify the occurrence of Event 1. We note that if Event 1 occurs, it can be very detrimental to the patient (e.g., cause death) even if it happens only once. Furthermore, among 7,518 patients who experienced Event 1 in our data, 7,408 (98.54%) experienced it only once within a time window ( $W = 30$  days). Therefore, in our analysis, we differentiate between no occurrence and at least one occurrence of Event 1 during each time window.

**3.3.2. Event 2 (UTP).** Pain relief is the first outcome measure in assessing the efficacy of pain medications (see, e.g., Teater (2015) and Smith et al. (2018)). When a patient is prescribed with a medication to address a pre-existing pain-inducing condition, but has post-supply visits due to the same condition, it is likely that the visits are due to undertreated pain (see, e.g., McPhillips-Tangum et al. (1998) and Xiao and Barber (2008)). As mentioned earlier, however, information on medications prescription is not typically linked to visits and corresponding diagnoses in claims data. As a result, identifying pain-inducing conditions that have triggered an opioid medication requires taking some extra steps. To this end, for each patient, we identify some Baseline Medical Conditions (BMCs, described below) and make use of them to define Event 2 (see Definition 2).

**Baseline Medical Conditions (BMCs).** In the absence of a surgical procedure, the purpose of prescribing painkillers is typically addressing existing acute or chronic pain conditions (for discussions related to the pathology of pain, see §3.2.3). Thus, for patients without surgical procedures, we characterize the BMCs using acute or chronic pain conditions pre-supply. In contrast, having a surgical procedure pre-supply would impact the characterization of the BMCs in two ways. First, a provider may prescribe pain medications because s/he is concerned about the underlying acute or chronic pain-inducing condition that prompted the surgery in the first place. Second, the provider could prescribe pain medications because of concerns related to pain conditions or complications arising after the surgical procedure in the pre-supply period. Therefore, for patients with surgeries, we characterize the BMCs based on a combination of (a) acute/chronic pre-surgical conditions, and (b) post-surgical pain-related conditions (PSPs) (see Appendix C.2 for the characterization of PSPs). Figure 2 illustrates the full steps taken to characterize the BMCs. Of note, in characterizing the BMCs, we assume that an opioid prescription is attributed to an acute condition first and then to a chronic condition. This is also consistent with the CDC’s recommendations (see, e.g., CDC (2016)).

DEFINITION 2 (EVENT 2). Let  $DX_n$  be a diagnosis among visits in time window  $n \geq 1$ . Then, we say that Event 2 has occurred in time window  $n$  if  $DX_n$  is among the baseline medical conditions:

$$\text{Event 2 (in window } n) = \begin{cases} 1, & \text{if } DX_n \in \text{BMCs,} \\ 0, & \text{if } DX_n \notin \text{BMCs.} \end{cases} \quad (2)$$

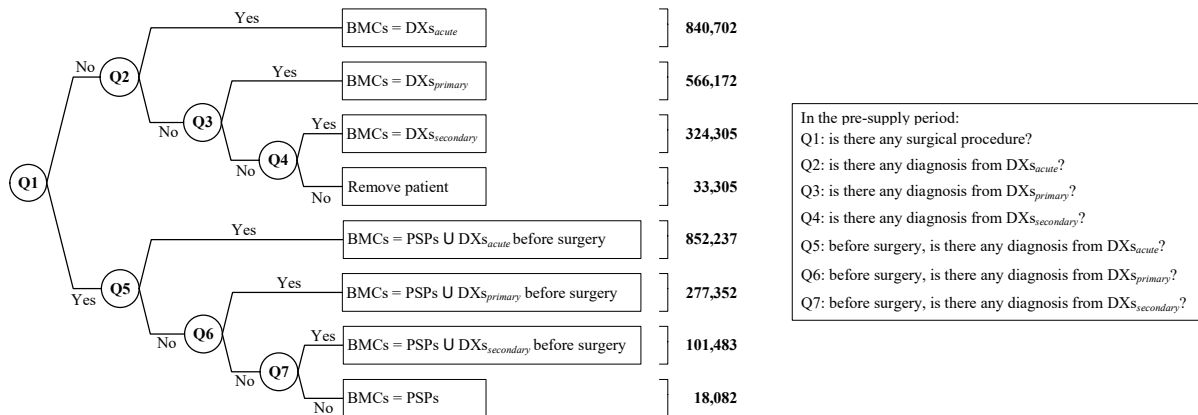
A value of 1 for Event 2 in a time window indicates that a baseline medical condition still exists in that window. Intuitively, compared to Event 2 = 1, Event 2 = 0 can indicate some potential benefits in using pain treatments.

REMARK 1 (IMPACT OF REPEATED VISITS). We establish Definition 2 and its impacts on the potential benefits of pain treatments based on the premise that repeated visits for a similar medical condition could indicate undertreated pain. However, we must also account for cases where a patient visits providers too often. Visiting a provider too often has two implications: it could inflate the occurrence of Event 2, and it might divulge the drug-seeking behavior of that patient, which is known to be a contributing factor to opioid use disorders (see, e.g., Grover et al. (2012)). Therefore, in addition to the independent variables described in §3.2, we adjust for another variable in our analysis: the total number of visits in each time window.<sup>4</sup>

## 4. The Analytical Setting

We use longitudinal machine learning algorithms to predict the risks of Events 1 and 2 with high accuracy for each individual. We then feed the best machine learning algorithm to an optimization

<sup>4</sup>Based on our data, the average (s.d.) of the number of visits in each window ( $W = 30$  days) is 1.89 (2.59).



**Figure 2 Characterization of the Baseline Medical Conditions (BMCs)**

Notes.  $DXs_{acute}$ ,  $DXs_{primary}$ ,  $DXs_{secondary}$ , and PSPs: sets of diagnoses codes for acute pain, chronic primary pain, chronic secondary pain, and post-surgical pain-related conditions, respectively. The value at the end of each branch represents the number of unique patients under that category based on the CCAE data.

model that allows us to determine cost-effective pain treatments based on each patient’s characteristics. We note that the black-box nature of the machine learning methods, along with the curse of dimensionality in the optimization model (due to multiple treatment modes), could negatively impact the implementation and interpretability of our approach for use in practice. Therefore, we propose two heuristics that are easy to adopt in a decision-support system. Later, in §5, we will compare the performance of these algorithmic-based heuristics with the human-based practices and show better cost-effective results under our proposed approaches.

#### 4.1. Longitudinal Machine Learning Algorithms

Most Machine Learning (ML) algorithms are used for *cross-sectional* studies, where subjects’ (e.g., patients’) information is recorded only at one point in time. However, our data is *longitudinal* in that patients are monitored and treated over time, and thus, their information, variables, and measurements dynamically change. Thus, we make use of ML methods that are suitable for longitudinal settings. *Generalized Estimating Equations Logistic Regression* (GEE Logit) is the first method that we use. GEE Logit takes advantage of the simplistic nature of logistic regression, while accounting for the longitudinal aspect of the data (see, e.g., Wilson and Lorenz (2015)). The second method we use is based on tree ensembles for longitudinal data (see, e.g., Capitaine et al. (2019) and Mišić (2020)). In particular, we adopt *Historical Random Forest* (HRF) (Sexton 2018). The third and fourth methods we use are based on Artificial Neural Networks. In particular, we make use of a *Recurrent Neural Network* (RNN) (Rumelhart et al. 1986) as well as a *Long Short-Term Memory* (LSTM) (Hochreiter and Schmidhuber 1997). Unlike GEE Logit, the other methods require tuning hyperparameters. In Table 5, we describe these parameters. For HRF, we have a tuple  $(t, n, v)$  which forms a three-dimensional grid of candidate values with  $5 \times 4 \times 7 = 140$  different combinations. For RNN and LSTM, we have a tuple  $(h, r, \gamma, w)$  which forms a three-dimensional

grid of candidate values with  $5^3 = 125$  different combinations. As noted by Hastie et al. (2009), in neural-network-based methods, when the number of epochs (denoted by  $r$ ) increases to infinity, the learning rate (denoted by  $\gamma$ ) decreases to zero. To reflect on this, we set  $\gamma = 10^{-(r-10)/10}$  (e.g., for  $r = 20$ ,  $\gamma = 0.1$ ), and do not consider variations for  $\gamma$  separately.

**4.1.1. Comparison of ML Methods under Various Hyperparameters.** After data pre-processing (see Appendix D), we apply our longitudinal ML methods to our data and compare their performance. We do so by measuring the area under the curve (AUC) under each method. The AUC values are, in turn, calculated using two validation mechanisms: 10-fold cross-validation (CV) and out-of-sample validation (OOS). Regarding CV, we use a 10-fold CV approach: we train the ML method based on 90% of randomly selected patients, and test it on the remaining 10%. We then repeat this procedure 10 times, and report the average AUC across these 10 folds. Regarding OOS, we create our train/test data sets based on the year in which a patient’s information was initiated. Specifically, we first create a training set from patients whose first encounter are recorded in 2008-2009, and then test the trained method on patients whose first encounters are in 2010. Afterwards, we create another training set based on 2008, and test the trained method on patients whose first encounter are recorded in 2009-2010. We then report the average AUC across these two iterations. In Table 6, we show the best hyperparameters—resulting in the highest AUC. In Figure 3, we depict the AUC values obtained from our ML methods under their best hyperparameters. Based on these results, we choose RNN with  $(h, r, \gamma, w) = (50, 20, 10^{-1}, 10^{-3})$  as the ML method with the best performance. In §4.2, we feed this trained RNN to an optimization model to determine the best multi-modal pain treatments.

**Table 5 Hyperparameters for our ML algorithms**

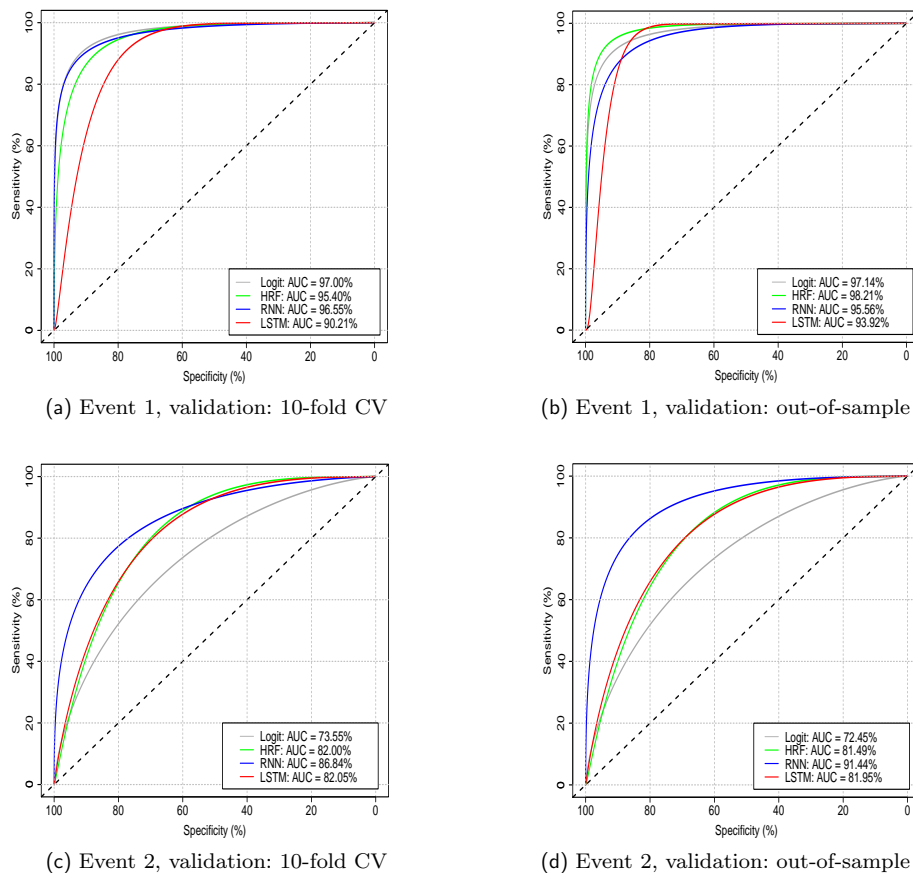
Method	Parameter	Candidate value(s)
HRF	$t$ : number trees to grow	(100, 200, 500, 800, 1,000) <sup>‡</sup>
	$n$ : minimum node size	(2, 5, 8, 10) <sup>‡</sup>
	$v$ : number of variable candidates to sample at each split	(1, 2, 3, 4, 5, 8, 12) <sup>§</sup>
RNN/LSTM <sup>†</sup>	$h$ : hidden dimension (size of the hidden layer)	(10, 20, 50, 100, 200) <sup>††</sup>
	$r$ : number of epochs (iterations over a training data)	(5, 10, 20, 50, 100) <sup>††</sup>
	$\gamma$ : learning rate of the algorithm	$\gamma = 10^{-(r-10)/10}$
	$w$ : learning rate (weight) decay	( $10^{-1}$ , $10^{-2}$ , $10^{-3}$ , $10^{-4}$ , $10^{-5}$ ) <sup>‡</sup>

<sup>†</sup>The same hyperparameters are used for these two methods. <sup>‡</sup>Candidate values are set based on values in Hastie et al. (2009). <sup>§</sup>We use  $\lceil 1.5^i(x/3) \rceil$  for  $i = -4, \dots, 2$  and  $x$ : # independent variables (see, e.g., Bertsimas et al. (2016)). <sup>††</sup>Candidate values are set based on those recommended by Choi et al. (2016).

**Table 6 Best hyperparameters for each ML algorithm under different validation mechanisms<sup>†</sup>**

Event	Method	Validation	
		10-fold cross validation	Out-of-sample
OUD	GEE Logit	—	—
	HRF	$(t^*, n^*, v^*) = (200, 5, 2)$	$(t^*, n^*, v^*) = (500, 8, 2)$
	RNN	$(h^*, r^*, \gamma^*, w^*) = (50, 20, 10^{-1}, 10^{-3})$	$(h^*, r^*, \gamma^*, w^*) = (50, 10, 1, 10^{-3})$
	LSTM	$(h^*, r^*, \gamma^*, w^*) = (100, 20, 10^{-1}, 10^{-3})$	$(h^*, r^*, \gamma^*, w^*) = (20, 50, 10^{-4}, 10^{-3})$
UTP	GEE Logit	—	—
	HRF	$(t^*, n^*, v^*) = (500, 2, 2)$	$(t^*, n^*, v^*) = (500, 2, 2)$
	RNN	$(h^*, r^*, \gamma^*, w^*) = (50, 20, 10^{-1}, 10^{-3})$	$(h^*, r^*, \gamma^*, w^*) = (100, 20, 10^{-1}, 10^{-5})$
	LSTM	$(h^*, r^*, \gamma^*, w^*) = (100, 10, 1, 10^{-3})$	$(h^*, r^*, \gamma^*, w^*) = (50, 10, 1, 10^{-5})$

<sup>†</sup>ROC curves under the best hyperparameters are shown in Figure 3.



**Figure 3 (Color online) Receiver operating characteristics curves under the best hyperparameters**

*Notes.* Logit: generalized estimating equations logistic regression, HRF: historical random forest, RNN: recurrent neural network, LSTM: long short-term memory.

## 4.2. Optimization Model

We aim to determine multi-modal pain treatment plans that are most cost-effective. To this end, for any treatment plan, we measure the *net monetary benefit* (see, e.g., Drummond et al. (2015)):

$$\text{Net monetary benefit} = \text{WTP} * \text{QALY} - \text{COST}, \quad (3)$$

where QALY is the *quality-adjusted life years* that a patient can accrue, WTP is the willingness to pay for an additional QALY, and COST is the total cost of care (e.g., reimbursement, co-payment, etc.).

Let  $\mathcal{A} = \mathcal{A}^{os} \times \mathcal{A}^{nos} \times \mathcal{A}^{ds} \times \mathcal{A}^{nph}$  be the set of all feasible actions representing all possible multi-modal treatment options ( $\times$ : the Cartesian product).<sup>5</sup>  $\mathcal{A}^{os} = [0, 100]$  MME and  $\mathcal{A}^{nos} = [0, 700]$  mg are feasible intervals for the opioid and non-opioid strengths, respectively.  $\mathcal{A}^{ds} = \{0, 1, \dots, 30\}$  days is the feasible set for the duration of supply of medications, and  $\mathcal{A}^{nph} =$

<sup>5</sup> Summary of all notations is provided in Appendix A.

$\{0 : \text{No non-pharmacologic treatment}, 1 : \text{Use non-pharmacologic treatment}\}$  is the set of possible actions with respect to non-pharmacologic treatments.<sup>6</sup> Also, let  $\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_T) \in \mathcal{A}^T$  be a sequence of multi-modal treatments (e.g., doses of opioid and non-opioid medications, their duration of supply, and use of non-pharmacologic treatments) taken throughout the time horizon  $T$ , and  $\mathbf{H}_t = (\mathbf{V}_1, \mathbf{a}_1, \dots, \mathbf{V}_{t-1}, \mathbf{a}_{t-1}, \mathbf{V}_t)$  be the history of the patient's covariates up to time window  $t$  (denoted by  $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_t$ ) as well as all the prior treatments. For Event  $k \in \{1, 2\}$ , we measure the cost and QALY accrued over the time horizon by:

$$\text{COST}_k(\mathbf{a}) = \sum_{t=1}^T \beta^{t-1} P_k(\mathbf{a}_t | \mathbf{H}_t) C_k, \quad (4a)$$

$$\text{QALY}_k(\mathbf{a}) = \sum_{t=1}^T \beta^{t-1} r_k(\mathbf{a}_t | \mathbf{H}_t), \quad (4b)$$

where  $\beta \in [0, 1)$  is a discount factor that allows us to prioritize the outcomes in the current period over those in the future,  $P_k(\mathbf{a}_t | \mathbf{H}_t)$  is the probability of experiencing Event  $k$  (obtained by our trained RNN) when the patient's history is  $\mathbf{H}_t$  and treatment is  $\mathbf{a}_t$ ,  $C_k$  is the total cost accrued due to an occurrence of Event  $k$  (including all payments to providers), and  $r_k(\mathbf{a}_t | \mathbf{H}_t)$  is the immediate QALY gained under  $\mathbf{a}_t$  and  $\mathbf{H}_t$ . We denote by  $q_{k0}$  ( $q_{k1}$ ) the monthly quality-of-life (qol) score for a patient who does not (does) experience Event  $k$ , where  $0 \leq q_{k1} \leq q_{k0}$ . Using this notation, the immediate QALY is measured as:

$$\begin{aligned} r_k(\mathbf{a}_t | \mathbf{H}_t) &= (1 - P_k(\mathbf{a}_t | \mathbf{H}_t)) \times q_{k0} + P_k(\mathbf{a}_t | \mathbf{H}_t) \times q_{k1} \\ &= q_{k0} - (q_{k0} - q_{k1}) \times P_k(\mathbf{a}_t | \mathbf{H}_t) \quad \text{for } t \leq T, \end{aligned} \quad (5)$$

where the first (second) part of summation on the first line indicates the expected QALY for a patient who is not experiencing (experiencing) Event  $k$ .

Using Equations (3)-(5), we measure the net benefit for Event  $k \in \{1, 2\}$  as:

$$\text{NB}_k(\mathbf{a}) = \text{WTP} \times \text{QALY}_k(\mathbf{a}) - \text{COST}_k(\mathbf{a}) \quad (6a)$$

$$= \sum_{t=1}^T \beta^{t-1} \left[ q_{k0} \times \text{WTP} - P_k(\mathbf{a}_t) \left( (q_{k0} - q_{k1}) \times \text{WTP} + C_k \right) \right], \quad (6b)$$

where, for notational brevity, we suppress the dependency of  $P_k(\cdot)$  and subsequent notations on  $\mathbf{H}_t$ .

<sup>6</sup> 100 MME and 700 mg are the 95<sup>th</sup>-percentile of values in our data for opioid and non-opioid strengths, respectively. Furthermore, the maximum duration of supply is equal to the length of a time window. Also, among patients who ever used non-pharmacologic treatments, 83.08% used either physical therapy or chiropractic. Therefore, to reflect on whether any non-pharmacologic option should be used, we consider a binary variable. However, we do not differentiate treatments based on the specific type of non-pharmacologic treatment, since the majority of non-pharmacologic treatment use relates to either physical therapy or chiropractic.



Using (6a)-(6b), we aim to find, for each patient, the treatment plan that maximizes the weighted average net benefit:

$$\max_{\mathbf{a} \in \mathcal{A}^T} \left\{ \text{NB}(\mathbf{a}) = \sum_{k=1}^2 w_k \text{NB}_k(\mathbf{a}) \right\} \quad \text{for } w_1, w_2 \in [0, 1] \text{ s.t. } w_1 + w_2 = 1, \quad (7)$$

where  $w_k$  is the weight assigned to Event  $k$ .

**4.2.1. Heuristic Solution Methods.** In finding optimal multi-modal pain treatments, we note that the feasible values for the opioid and non-opioid strengths are continuous. Even considering a grid-based uniform discretization (e.g.,  $\{0, 10, \dots, 100\}$  MME and  $\{0, 50, \dots, 700\}$  mg for opioid and non-opioid strengths, respectively) results in  $11 \times 15 \times 31 \times 2 = 10,230$  different treatment combinations to be explored within the feasible action space  $\mathcal{A}$  for each time window causing a curse of dimensionality. To address this curse, we consider two heuristic approaches: *myopic* and *rolling-horizon*. In the myopic approach, we select the treatment resulting in the highest average net benefit in each window. That is, we only value the current outcomes and do not consider the impact of current decisions on future outcomes. In the rolling-horizon approach, we determine the sequence of optimal treatments for a sub-horizon that is computationally tractable (in our experiments, we consider a sub-horizon of 3 time windows), and move this sub-horizon one time-window forward as we go. Compared to the myopic policy, this accounts for the impact of current decisions on both current and future outcomes (see Table 7 for a summary of these heuristic approaches).

## 5. Numerical Experiments

To perform our main numerical experiments, we make use of the parameter values shown in Table 8. In §5.2, we perform extensive robustness checks on these parameters, and test the validity of our main findings by varying them.

### 5.1. Comparison with the Human-Based Guidelines and Practice

**5.1.1. Cost-Effectiveness of Algorithmic-Based Treatment Policies.** In this section, we compare the cost-effectiveness of our algorithmic-based treatment policies with those of the CDC guidelines as well as the human-based approaches followed in practice by physician experts, as evidenced from our data. To this end, we develop a simulation model where we first estimate distributions of multi-modal pain treatments from our data. Then, we adjust these estimated distributions by the thresholds recommended by the CDC 2016 guidelines (further details are provided in Appendix E). For any WTP  $> 0$ , our treatment policy is said to be more cost-effective than the guidelines if (see, e.g., Drummond et al. (2015)):

$$\begin{aligned} \text{Incremental Cost-Effectiveness Ratio (ICER)} = \\ \frac{\text{Incremental Cost}}{\text{Incremental QALY}} = \frac{\text{Cost(our policy)} - \text{Cost(guidelines)}}{\text{QALY(our policy)} - \text{QALY(guidelines)}} \leq \text{WTP}. \end{aligned} \quad (8)$$

**Table 7** Summary of heuristic solution approaches

<b>Inputs:</b> trained RNN, action space $\mathcal{A}$ (grid-based), time horizon $T$ , sub-horizon $L < T$ , discount factor $\beta$ vectors of patient covariates $\mathbf{V}_1, \dots, \mathbf{V}_T$ , $\mathbf{H}_1 = (\mathbf{V}_1)$	
<b>Solution Approach: RNN-Myopic</b>	
1	for $t = 1, \dots, T$
2	for each $\mathbf{a}_t \in \mathcal{A}$
3	measure the risk of experiencing Event 1 (OUD) and Event 2 (UTP) by the trained RNN
4	measure the weighted avg. net benefit $\text{NB}(\mathbf{a}_t   \mathbf{H}_t)$ by Equations (6a)-(7)
5	select treatment $\mathbf{a}_t^* = \arg \max_{\mathbf{a}_t \in \mathcal{A}} \{\text{NB}(\mathbf{a}_t   \mathbf{H}_t)\}$
6	if $t \leq T - 1$
7	update history $\mathbf{H}_{t+1} = (\mathbf{V}_1, \mathbf{a}_1^*, \dots, \mathbf{V}_t, \mathbf{a}_t^*, \mathbf{V}_{t+1})$
<b>Solution Approach: RNN-Rolling-Horizon</b>	
1	for $t = 1, \dots, T$
2	for $l = t, \dots, t + L - 1$
3	for each $\mathbf{a}_l \in \mathcal{A}$
4	measure the risk of experiencing Event 1 (OUD) and Event 2 (UTP) by the trained RNN
5	measure the weighted avg. net benefit $\text{NB}(\mathbf{a}_l   \mathbf{H}_l)$ by Equations (6a)-(7)
6	select the sequence of treatments $(\mathbf{a}_t^*, \dots, \mathbf{a}_{t+L-1}^*) = \arg \max_{\mathbf{a}_t, \dots, \mathbf{a}_{t+L-1} \in \mathcal{A}} \{\sum_{l=t}^{t+L-1} \beta^{l-t} \text{NB}(\mathbf{a}_l   \mathbf{H}_l)\}$
7	if $t \leq T - 1$
8	update history $\mathbf{H}_{t+1} = (\mathbf{V}_1, \mathbf{a}_1^*, \dots, \mathbf{V}_t, \mathbf{a}_t^*, \mathbf{V}_{t+1})$
<b>Output:</b> treatments $\mathbf{a}_1^*, \dots, \mathbf{a}_T^*$	

**Table 8** Summary of baseline parameters

Parameter	Description
$T = 12$	Time horizon (12 months)
$\text{WTP} = 10,000$ (\$/QALY)	Willingness to pay
$q_{10} = q_{20} = 1/12 = 0.083$ yr <sup>†</sup>	Monthly qol score for not experiencing Events 1 or 2 (i.e., perfect health)
$q_{11} = 0.368/12 = 0.031$ yr <sup>†</sup>	Monthly qol score for experiencing Event 1 (see, e.g., De Maeyer et al. (2010))
$q_{21} = \begin{cases} 0.395/12 = 0.033$ yr (acute) \\ 0.470/12 = 0.039 yr (chronic)	Monthly qol score for experiencing Event 2 (for an acute and chronic pain pre-supply) (see, e.g., Katz (2002), Wu et al. (2003), Tüzün (2007), Ataoğlu et al. (2013), Taylor et al. (2013), Hadi et al. (2019) and references therein)
$C_1 = \$15,588.07$	Costs associated with an occurrence of Event 1 <sup>§</sup>
$C_2 = \begin{cases} \$6,172.86$ (acute) \\ \$2,714.90 (chronic)	Costs associated with an occurrence of Event 2 (for an acute and chronic pain) <sup>§</sup>
$w_1 = w_2 = 0.5$	Weights assigned to Events 1-2

<sup>†</sup>Yearly qol scores are converted to their monthly equivalents. <sup>‡</sup>We consider a same qol score for both genders (see, e.g., Giacomuzzi et al. (2005) and Domingo-Salvany et al. (2010)). <sup>§</sup>To estimate these values, we first obtain the average cost from our data, and then use the average U.S. healthcare inflation rate (USIC 2021) to prorate the costs from years 2008-2010 (time frame of our data) to year 2021.

We iterate our simulation 10,000 times to account for variations in dynamic risk covariates and treatments generated by the guidelines. The percentage of instances (out of 10,000) that satisfy (8) measures the cost-effectiveness probability (see, e.g., Fenwick et al. (2006)).

The CDC has recently (in 2022) put forth a new set of guidelines to rectify some of the issues attributed to the 2016 guidelines. In particular, they proposed to remove the recommended thresholds for the opioid dose and duration of supply (90 MME for opioid dose and a total of 7 (90) days for the duration of supply for acute (chronic) pain). However, some believe that the new guidelines have the potential for human misunderstanding or misapplication, especially since they still emphasize on the harms of opioids rather than the harms of poorly managed pain. In addition, some other experts are wary against removing the aforementioned thresholds, and believe that doing such could make it more difficult for physicians who have relied on these thresholds in their practices over the past six years (for more details, see CDC (2022) and NPR (2022)). To

gain some insights, we also compare our algorithmic-based approach with the CDC guidelines after incorporating the 2022 modifications (see Appendix E for more details).

For both the 2016 and 2022 guidelines, we also run our simulations under various opioid tapering rates. We do so because in practice an initial opioid dose is typically *tapered down* over time by a linear rate exogenously (see, e.g., CDC (2016) and FDA (2019)). For example, if an initial dose is 50 MME and the tapering rate is 20% per month, then the resulted dose in the second month will be  $50 \times 0.8 = 40$  MME. This is done with the purpose of balancing the risks of opioid-related disorders and opioid-withdrawal symptoms. To reflect on this, we simulate each set of guidelines under different tapering rates denoted by  $\theta \in \{0.05, 0.20, 0.50\}$  (VA.gov 2016).

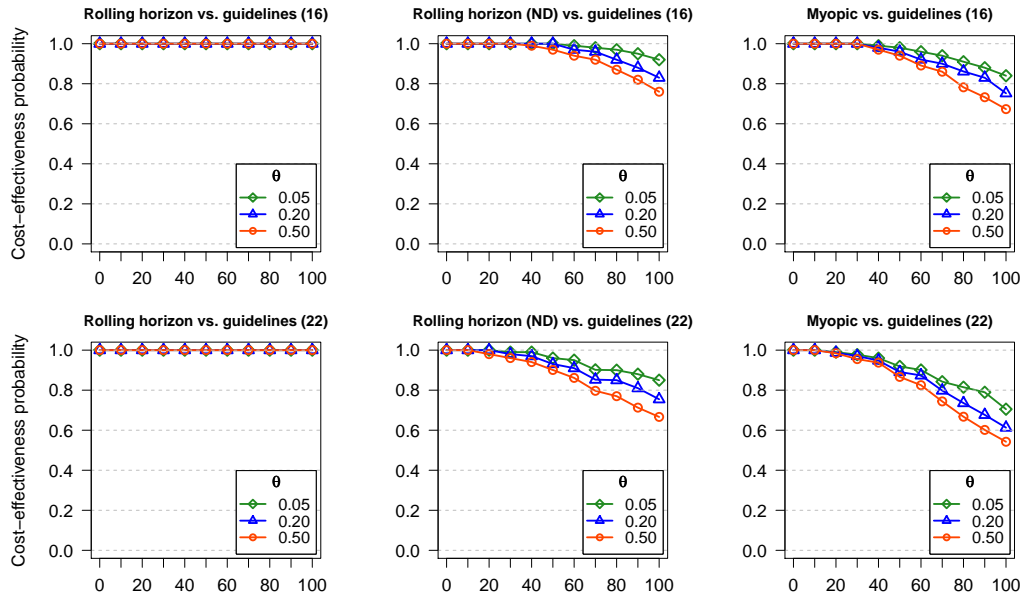
Finally, we note that, under the treatment policies we obtain, the duration of supply can be a fraction of a time window, which might cause a temporary discontinuation of medications. As indicated by FDA (2019), “*Health care professionals should not abruptly discontinue opioids in a patient who is physically dependent.*” To address this, we simulate another version of our treatment policies in which we disallow any discontinuation with respect to chronic pain management. That is, we fix the duration of supply in each window to be equal to 30 days. Of note, the supply is still equal to zero in a time window if opioid and non-opioid doses are both zero in that window.

Based on our results in Figure 4, we make the following observation.

OBSERVATION 1. (i) *Our algorithmic-based treatment policies are more cost-effective than both the CDC 2016 and 2022 guidelines, but this effectiveness typically reduces as WTP increases.* (ii) *The 2022 CDC guidelines are more cost-effective than the 2016 CDC guidelines.*

Observation 1(i) indicates that our proposed approach yields treatment policies that are more cost-effective than the CDC guidelines. This is to some extent expected, because our algorithmic-based treatment policies are obtained by maximizing the net monetary benefit (see Equation (3)). Our results, however, show that even when we disallow any discontinuation of medications for chronic pain management, our obtained policies are still more cost-effective than the CDC guidelines. However, this effectiveness drops as we increase WTP. For example, compared to the 2022 CDC guidelines (with a tapering rate  $\theta = 0.05$ ), the treatment policy we obtain with no discontinuation of supply is cost-effective in 96.03% and 85% of simulated instances under  $WTP = \$50K$  and  $\$100K$ , respectively.

Observation 1(ii) implies a better performance under the 2022 guidelines than the 2016 ones. For example, compared to the CDC 2016 guidelines (with a tapering rate  $\theta = 0.50$ ), our algorithmic-based treatments with no discontinuation of supply is cost-effective in 97% and 76% of simulated instances under  $WTP = \$50K$  and  $\$100K$ , respectively. However, these numbers under the 2022 CDC guidelines drop to 90.09% and 66.64%, respectively. This can be due to the fact that the 2022 guidelines have lifted the recommended thresholds on the opioid dose and days of supply, which could, in turn, help reduce instances of undertreated pain. As the public comment period for the



**Figure 4** Cost-effectiveness of our algorithmic-based policies compared to the human-based guidelines

*Notes.* x-axes represent WTP (\$1,000/QALY). Results are the avg. cost-effectiveness probabilities from simulation instances.  $\theta$  is the opioid dose-tapering rate. ND: no discontinuation of medications supply for chronic pain. 16 and 22 represent the CDC’s 2016 and 2022 medical guidelines, respectively. The lower the probability, the better the performance of the guidelines compared to our proposed policies.

first draft of the new guidelines recently ended in April 2022, the CDC will soon evaluate the final recommendations (NPR 2022). We hope that our findings here could provide further insights for healthcare policymakers in their efforts in improving medical guidelines based on clinical evidence.

Finally, our results show that the average (s.d.) improvement in the QALY and cost per patient per year due to using our treatment policy are 2.82 (2.24) days and \$461.46 (\$635.92), respectively. Given that a large number of patients are affected by opioid prescriptions each year, and that the opioid pandemic has been lingering for many years, we believe these per patient per year improvements offer significant advantages at the society level for policymakers to reconsider some of their recommendations.

**5.1.2. Multi-Modal Pain Treatments.** In the previous section, we reported the results for an average patient whose risk covariates could change throughout the time horizon. In §5.1.2-5.1.4, we aim to analyze medical outcomes based on patients’ heterogeneity. To this end, and to handle the sheer volume of covariates scenarios, we form cohorts of patients based on their characteristics pre-supply, such as age, gender, pain pathology, history of surgery or an inpatient admission, and history of behavioral factors (more details about these cohorts are provided in Appendix F.1).

For each cohort of patients, Figure 5 shows the multi-modal pain treatment plan we obtain from our rolling-horizon approach as well as those we observe physicians have followed based on our data. The results indicate that for any type of pain, existence of various heterogeneity factors, such as

age, gender, and history of behavioral factors, has hardly any impact on the intensity of treatments prescribed by physicians in the practice. This “one-size-fits-all” type of pain management was also suggested in the CDC 2016 guidelines, where no more than 50-90 MME for the opioid strength or no more than 7 (90) days of opioid supply for acute (chronic) pain were recommended (see, e.g., Dowell et al. (2016) and AZDHS (2017)). To the best of our knowledge, there is no evidence-based information on how these treatments should vary based on patients’ characteristics. Thus, we make the following observation from our results in Figure 5.

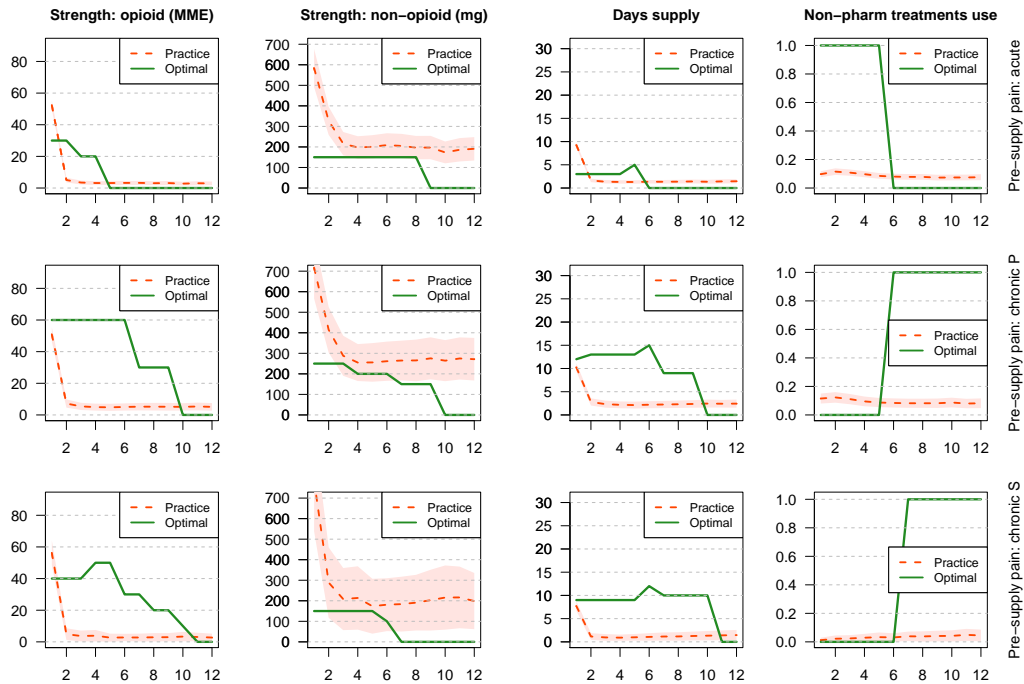
*OBSERVATION 2. Compared to the current human-based approach, our algorithmic-based treatment policies result in*

- (i) more (less) intensive opioid regimens for patients with chronic pain and no pre-supply history of behavioral factors (acute or chronic pain and a pre-supply history of behavioral factors),*
- (ii) more days of medications supply for females in their 40’s with either both history of surgery/inpatient admission and behavioral factors pre-supply (when having chronic pain) or neither of these histories (when having acute or chronic pain),*
- (iii) more days of medications supply for males in their 50’s with either no history of surgery or inpatient admission pre-supply (when they have chronic pain and a history of behavioral factors), or a history of surgery or inpatient admission pre-supply (when they have acute or chronic pain and no history of behavioral factors), and*
- (iv) more (less) intensive use of non-pharmacologic treatments (non-opioid regimens) across all cohorts of patients.*

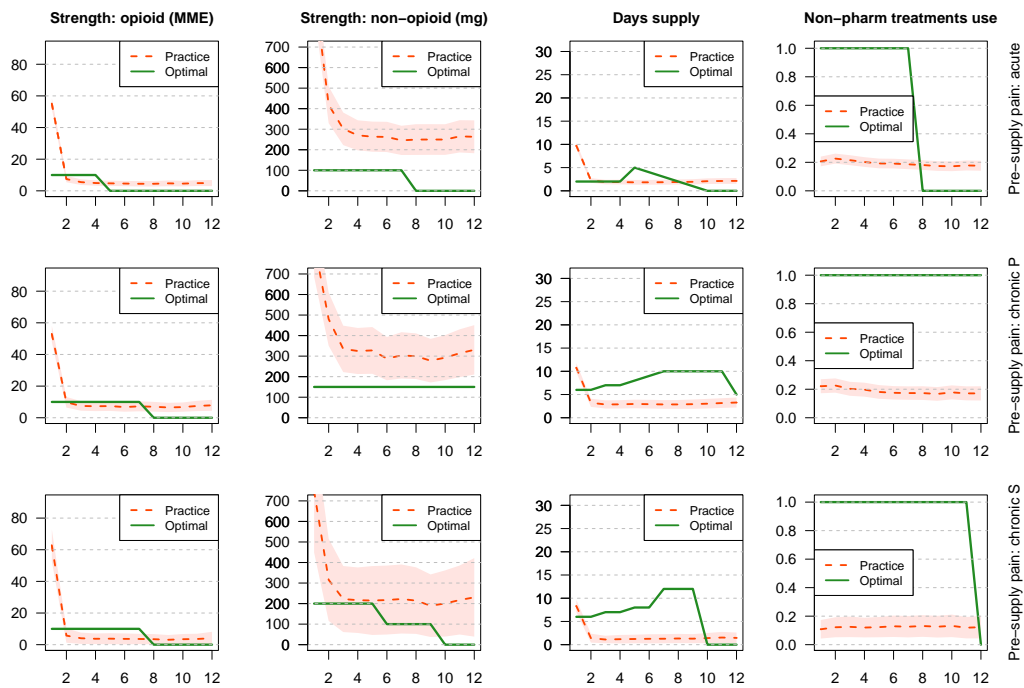
Observation 2(i) implies that patients with chronic pain and no history of behavioral factors pre-supply (e.g., alcohol consumption or mental health disorders) will be prescribed with a higher opioid dose under our treatment policies compared to the observed practice. Also, for females in their 40’s or those with a history of surgeries/inpatient admissions pre-supply, this higher dose will be prescribed more aggressively (i.e., the difference in dose is more significant).

As mentioned before, medical guidelines recommend an initial opioid dose to be tapered down exogenously (see, e.g., FDA (2019)). Compared to this approach, our algorithm-driven treatment policies account for various risk covariates to personalize this tapering behavior. This induces an endogenous dose tapering, which is often nonlinear in time. For example, among patients with acute/chronic pain and no history of behavioral factors pre-supply, our policies result in steeper tapering for the following cohorts: females in their 40’s and patients with a history of surgeries/inpatient admissions pre-supply.

Furthermore, our results indicate that, when a patient has a chronic pain pre-supply, the duration of supply should increase in the early stage of therapeutic course and then be reduced towards the end of time horizon. This finding holds for patients with different age, gender, behavioral factors, and surgery/inpatient admission records. The medical reason behind this finding is that



(a) Female, 40 years old, surgery/inpatient adm pre-supply: yes, history of behavioral factors pre-supply: no



(b) Female, 40 years old, surgery/inpatient adm pre-supply: yes, history of behavioral factors pre-supply: yes

**Figure 5 Multi-modal pain treatments obtained from our rolling-horizon policy and the practice**

Notes. x-axes represent time windows (month). Results from the rolling-horizon policy are obtained under  $WTP = \$10K$ . Shades represent 95% confidence intervals. Supply is zero when the opioid/non-opioid strengths are both zero. non-pharmacologic treatments: 1 (0) implies use (no use). Chronic P/S: chronic primary/secondary. Results for other cohorts are presented in Appendix F.2.

the increase in the duration of supply could balance the tapering down of medications, which, in turn, can mitigate the potential risks of opioid withdrawal symptoms.

Observation 2(iv) indicates that human-based approaches (in which a human-based judgment is used to assess the potential benefits and risks) typically prescribe more intensive non-opioid medications compared to our algorithmic-based treatments, and this is more aggressive for acute or chronic primary (compared to chronic secondary) pain pre-supply. In addition, this finding is more prominent among patients who either do not have a history of surgeries/inpatient admission pre-supply, or those who have such a history but have also shown a history of behavioral factors.

We also observe from our results that the rate of using non-pharmacologic treatments (NPHTs) under human-based approaches is typically higher for patients with chronic pain and those with a history of behavioral factors or surgeries/inpatient admissions pre-supply compared to other patients. Overall, however, our algorithm-driven treatment policies often make use of NPHTs at higher rates than the human-based approaches in the practice.<sup>7</sup> In particular, we observe that for patients with an acute pain pre-supply, NPHTs should be used for a few months after the onset of pain. In addition, this period should be extended for patients with a history of behavioral factors or a surgery/inpatient admission pre-supply. Furthermore, when there is a history of behavioral factors, NPHTs should be used over the whole time horizon.<sup>8</sup> However, when there is no such history, NPHTs should be used towards the later stage of time horizon, because this is the time when the dose of pharmacologic treatments (PHTs) is already tapered down. This implies a *complementary* relationship between NPHTs and PHTs, which, in turn, substantiates the importance of a comprehensive and orchestrated multi-modal pain management plan.

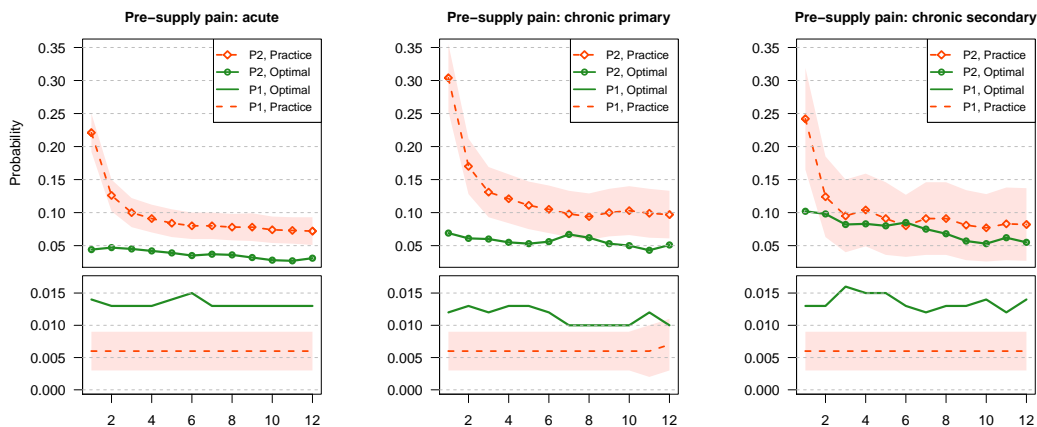
**5.1.3. Risks of Events 1 and 2.** We now compare the incidences of Events 1 and 2 under our algorithm-driven treatment policies and the human-based approaches observed from the medical practice. For the former, we follow the optimal treatments (discussed in §5.1.2) in our trained RNN and obtain the risks of Events 1 and 2. For the latter, we monitor the incidence of Events 1 and 2 from our data, and create empirical 95% confidence intervals. Our results are presented in Figure 6. From this figure, we make the following:

*OBSERVATION 3. Compared to the current human-based practice, our treatment policies result in a slightly higher risk of Event 1 but considerably lower risk of Event 2.*

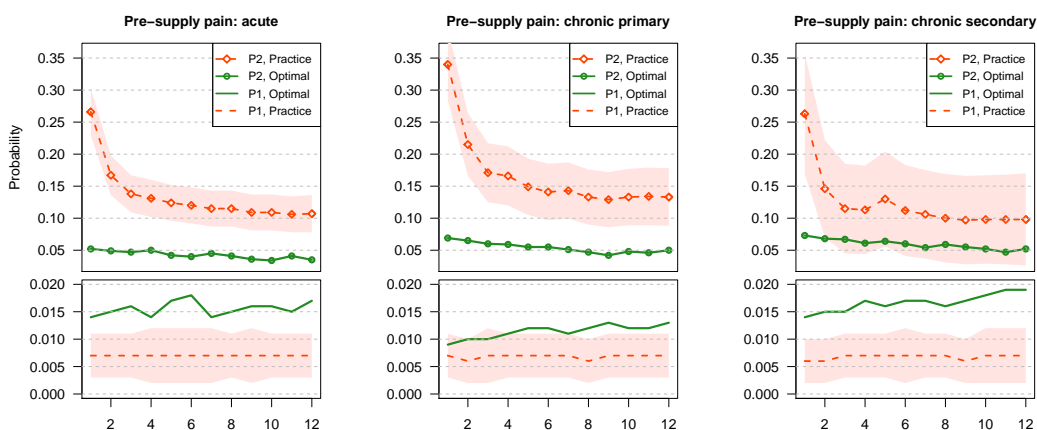
Observation 3 indicates that, under our treatment policies, the risk of Event 1 is slightly higher than that observed from the practice. We also note that this gap is more prominent for patients

<sup>7</sup> One reason for the low utilization of NPHTs in the practice compared to our treatment policies could be the lack of coverage for some of these services by insurance companies (see, e.g., Boloori et al. (2020b)). Nevertheless, we hope that our findings here could contribute to the various efforts in adopting alternative treatments in pain management (see, e.g., Goertz and George (2018) and Johns Hopkins (2018)).

<sup>8</sup> One exception is when there is no history of surgery or inpatient admission pre-supply, which makes the use of NPHTs over the time horizon less frequent.



(a) Female, 40 years old, surgery/inpatient adm pre-supply: yes, history behavioral factors pre-supply: no



(b) Female, 40 years old, surgery/inpatient adm pre-supply: yes, history behavioral factors pre-supply: yes

**Figure 6** Likelihoods of incidence of Events 1 and 2 obtained from our rolling-horizon policy and the practice

Notes. P1 (P2): risk of Event 1 (2). x-axes represent time windows (month). Results from the rolling-horizon policy are obtained under WTP = \$10K. Shades represent 95% confidence intervals. Results for other cohorts are presented in Appendix F.3.

with (1) chronic primary pain who also have a record of surgeries/inpatient admission pre-supply, or (2) no history of behavioral factors pre-supply who are either females in their 40's with no record of surgeries/inpatient admission pre-supply or males in their 50's. However, unlike the risk of Event 1, the risk of Event 2 resulted from our treatment policies is considerably lower than that observed in the practice. The gap between the risk of Event 2 under the treatment policies we obtain and those in the practice is particularly large among patients with (1) chronic primary pain who are either males in their 50's with a history of surgeries/inpatient admission and no history of behavioral factors pre-supply or females in their 40's, and (2) a history of behavioral factors pre-supply who also have a record of surgeries/inpatient admission pre-supply. Overall, these results indicate that our treatment policies are able to use patient-specific covariates to *personalize* the balance between the risks of Events 1 and 2. On average, however, they improve performance



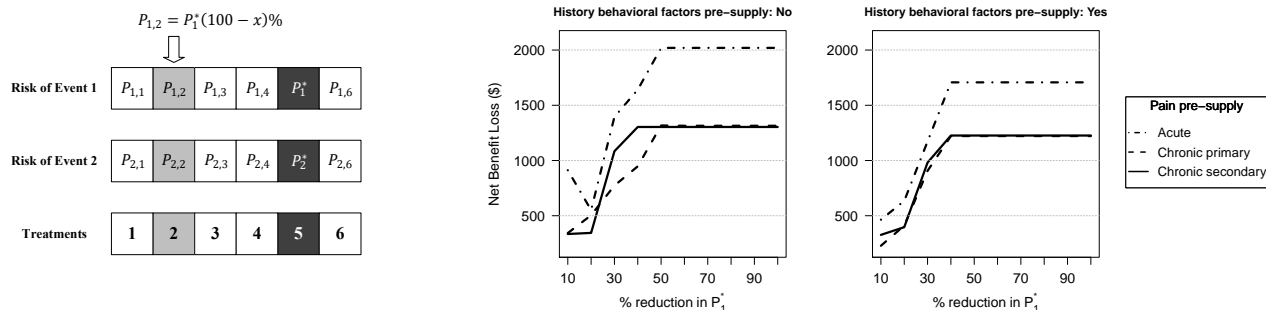
compared to the observed practice by slightly increasing the risk of Event 1 while significantly decreasing that of Event 2.

To gain deeper insights into changes that our algorithm-driven treatment policies would cause, if adopted in practice, we also investigate the relative risks of experiencing Event 2 compared to Event 1. That is, under both our treatment policies and the observed practice, and for different cohorts of patients, we measure the number of patients that would have to experience undertreated pain for one similar patient to experience opioid-related disorders at any time during the therapeutic course. This sheds light on a quantity known as “number needed to undertreat,” which the American Academy of Pain Medicine deems crucial for physicians in their daily practice (Carr 2016). Based on the results in Table 9, we observe that the relative risk of Event 2 to Event 1 in the observed practice is typically higher among (1) patients with chronic primary pain pre-supply, (2) females in their 40’s who have either both history of surgeries/inpatient admissions and behavioral factors or neither of these records pre-supply, and (3) patients with a history of surgeries/inpatient admissions and behavioral factors pre-supply. In contrast, under our treatment policies, the relative risk of Event 2 to Event 1 is higher among (1) patients with chronic primary pain pre-supply, (2) males in their 50’s, and (3) patients with history of surgeries/inpatient admissions who do not have a history of behavioral factors pre-supply.

We also note that the main goal of the CDC guidelines has been centered around minimizing the risk of Event 1. Our results discussed thus far indicate that, by and large, this is not a good

**Table 9 Average (S.D.) of relative risk of Event 2 to Event 1 over the time horizon**

		Cohort			Policy	
Gender	Age	Surgery/inpatient admission pre-supply	History of behavioral factors pre-supply	Pain pathology pre-supply	Practice	Optimal
Female	40’s	Yes	No	Acute	16.07 (7.01)	2.77 (0.51)
				Chronic primary	21.10 (10.04)	5.03 (0.89)
				Chronic secondary	17.24 (7.59)	5.56 (1.29)
		No	Yes	Acute	19.13 (6.43)	2.77 (0.57)
				Chronic primary	24.35 (8.97)	4.85 (1.30)
				Chronic secondary	18.58 (8.46)	3.63 (0.81)
	50’s	Yes	No	Acute	7.39 (4.12)	2.24 (0.56)
				Chronic primary	13.50 (7.18)	3.99 (0.89)
				Chronic secondary	10.76 (4.46)	4.41 (1.77)
		No	Yes	Acute	7.75 (4.11)	7.62 (1.96)
				Chronic primary	14.31 (6.87)	7.27 (1.45)
				Chronic secondary	12.22 (4.64)	7.57 (2.10)
Male	40’s	Yes	No	Acute	18.32 (10.16)	3.31 (0.66)
				Chronic primary	21.79 (11.39)	6.79 (1.17)
				Chronic secondary	18.50 (7.37)	8.36 (2.09)
		No	Yes	Acute	19.55 (8.63)	6.10 (1.87)
				Chronic primary	23.32 (10.25)	7.96 (2.51)
				Chronic secondary	17.55 (5.09)	6.33 (1.67)
	50’s	Yes	No	Acute	4.97 (3.48)	2.47 (0.86)
				Chronic primary	9.34 (4.44)	3.90 (1.22)
				Chronic secondary	9.38 (2.50)	5.54 (1.99)
		No	Yes	Acute	8.19 (4.19)	7.44 (1.89)
				Chronic primary	12.81 (6.19)	9.74 (1.82)
				Chronic secondary	17.83 (3.91)	9.80 (2.30)



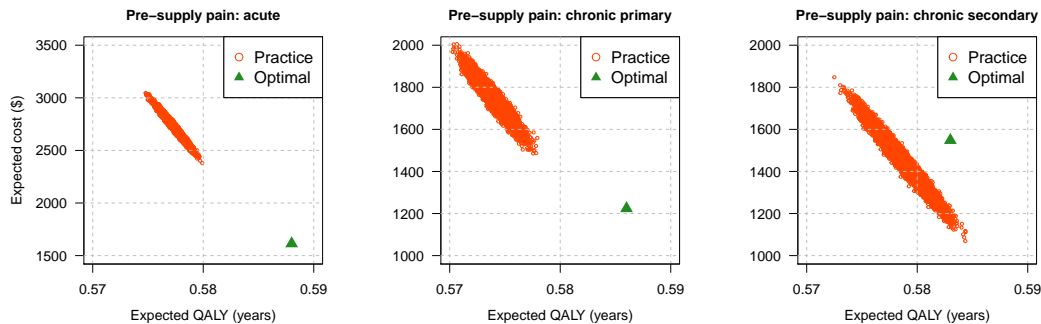
**Figure 7** Impact of the reduction in the optimal risk of Event 1 ( $P_1^*$ )

*Notes.* (Left) A schematic illustration on how  $x\%$  reduction in  $P_1^*$  alters the treatment (and hence the risks of Events 1-2). The net benefit loss is the difference between the net benefits when following the altered (instead of optimal) treatments throughout the time horizon. (Right) Net benefit loss. Results from the rolling-horizon policy are obtained for WTP = \$10K and 50 year-old males with a history of surgeries/inpatient admissions pre-supply. Results for other cohorts are presented in Appendix F.4.

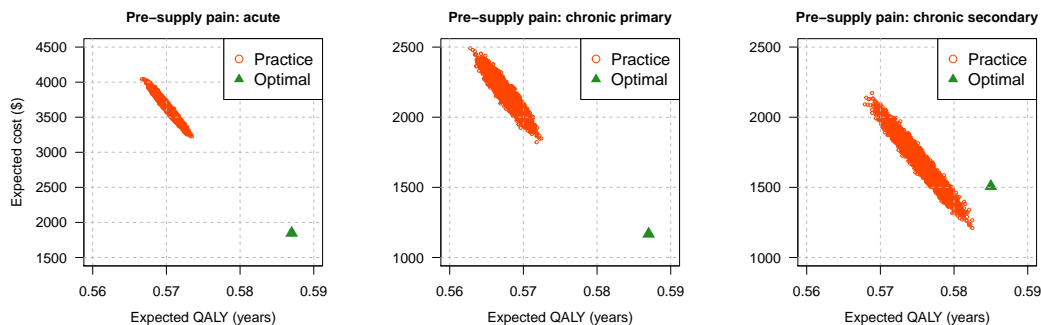
objective to follow: what is needed is a careful and personalized balance between risks of Events 1 and 2. However, to gain a better understanding of the importance of this, we consider a hypothetical policy in which the risk of Event 1 is fully reduced to 0.<sup>9</sup> Let  $P_1^*(t)$  be the risk of Event 1 in window  $t$  resulted from the rolling-horizon policy and  $P_1(t) = P_1^*(t)(100 - x)\%$  be the risk resulted from an  $x\%$  reduction in  $P_1^*(t)$ . It should be note that any such reduction in the risk of Event 1 affects the risk of Event 2 (see Figure 7). To measure the overall impact, we let  $NB^*$  and  $NB$  be the corresponding total monetary net benefit when using these risks over the time horizon, respectively. In Figure 7, we show how the total net benefit loss,  $NB^* - NB$ , changes based on variations in  $x$ . If we follow treatments resulting in no risk of Event 1 over the therapeutic course (i.e., a 100% reduction in  $P_1^*(t), \forall t \leq T$ ), we find that, for patients with a history of behavioral factors pre-supply, the society would lose up to \$1,700 of net benefit per patient (this holds across patients with different age, gender, and surgery records). Furthermore, when there is no history of behavioral factors pre-supply, the net benefit loss would be up to \$2,000 (\$1,300) per patient in the presence (absence) of surgeries/inpatient admission pre-supply. We also observe that the net benefit loss is typically higher for patients with acute (compared to chronic) pain pre-supply, although the former has a lower risk of Event 2 compared to the latter. This can be due to the fact that the cost of experiencing Event 2 is higher when the pain-inducing condition is acute (compared to chronic) (see Table 8).

Put together, our results in this section indicate that the CDC's aim of solely reducing incidence of opioid-related disorders can be harmful. Instead, we find that following algorithmic-based policies that make a personalize balance between reducing the risk of opioid-related disorders and increasing that of under-treated pain can go a long way.

<sup>9</sup> As Figure 6 shows, a drop in the risk of Event 1 would coincide with a much higher increase in the risk of Event 2. Therefore, the drop in the cost of Event 1 incidence could come at the expense of more frequent visits due to higher incidence of Event 2. This can negatively impact the net benefit gained.



(a) Female, 40 years old, surgery/inpatient adm pre-supply: yes, history behavioral factors pre-supply: no



(b) Female, 40 years old, surgery/inpatient adm pre-supply: yes, history behavioral factors pre-supply: yes

**Figure 8 QALY and cost gained per patient from our rolling-horizon policy and the practice**

*Notes.* Results from the rolling-horizon policy are obtained under  $WTP = \$10K$ . To account for the variability of the risks of Events 1 and 2 under the practice, we randomly select 10,000 pairs of these risks from the corresponding confidence intervals (shown in Figure 6) and report the (QALY, Cost) pair here. Results for other cohorts are presented in Appendix F.5.

**5.1.4. Who Benefits the Most?** We now investigate the type of patients that would benefit most from our treatment policies compared to the practice of physicians. To this end, we make use of the risks of Events 1 and 2 (illustrated in Figure 6) to measure the QALY and cost gained. Based on the results in Figure 8, we make the following observation.

*OBSERVATION 4.* *By following our algorithmic-based treatment policies, patients with the following conditions pre-supply would benefit most in terms of both QALY and cost: acute or chronic primary pain and a history of behavioral factors or surgery/inpatient admission. This finding holds across patients with different age or gender.*

As discussed earlier under Observation 3, the risk of Event 2 (1) is lower (higher) under our treatment policies compared to the observed practice of physicians. Moreover, across these policies, the difference in the risk of Event 2 (1) is more (less) prominent for patients with acute or chronic primary pain compared to those with chronic secondary pain pre-supply. A similar result holds for patients with a record of surgery/inpatient admission or behavioral factors pre-supply.

## 5.2. Robustness Checks

We conduct a variety of robustness checks on our baseline parameters (Table 8) to measure the sensitivity of our results to various potential misspecifications. In what follows, we focus on presenting our robustness checks with respect to our cost-effectiveness results. We do so because many of our main results are driven by the fact that our framework allows obtaining treatment policies that are not only more personalized, and hence, effective in improving patient outcomes (improving QALY), but are also less costly than CDC guidelines or the observed practice of human-experts. Thus, if our sensitivity results indicate that our treatment policies remain cost-effective even when we alter the estimated parameters, it can give us further confidence about the validity of our main findings.

**qol scores.** Compared to the values under the baseline setting, we consider alternative scenarios where the qol scores for Events 1 and 2 are perturbed (increased or decreased). Based on the results presented in Table 10, we observe that the treatment policies we obtain remain cost-effective even when we alter the qol scores. The cost-effectiveness of our policies compared to the guidelines in particular improve under two circumstances: when the qol score for experiencing Event 1 (2) increases (decreases), which is the case when the severity of opioid-related disorders decreases (e.g., opioid dependence compared to abuse/poisoning) and/or when the level of undertreated pain increases.

**Costs.** Similarly, we consider different scenarios for the cost of experiencing Events 1 and 2 by altering the estimated values we used in our baseline setting. From the results presented in Table 10, we find that our treatment policies remain cost-effective compared to the CDC guideline even when the estimated cost parameters are perturbed. Moreover, decreasing (increasing) the cost of experiencing Event 1 (2) further improves the performance of our proposed policies compared to the guidelines. This is consistent with our observation regarding the qol scores, since lower qol scores indicate worse health conditions, which typically require more medical services/procedures (thus, resulting in higher costs).

**Weights.** Under our baseline setting, we assigned equal weights to the outcomes related to Events 1 and 2. Compared to this, we consider alternative scenarios where we change the assigned weights. Our results presented in Table 10 show that when the weight assigned to Event 1 increases (e.g.,  $w_1 = 0.75$ ,  $w_2 = 0.25$ ), our policies are still more cost-effective compared to the guidelines. However, when we only account for opioid-related disorders (i.e.,  $w_1 = 1.0$ ,  $w_2 = 0.0$ ), the performance of our policies drastically diminish compared to the guidelines. This is yet another indication that the CDC guidelines mainly focused on avoiding opioid-related disorders, while largely overlooking other important outcomes such as pain remaining undertreated.

Overall, our results show that the performance of our algorithm-driven treatment policies (against the guidelines) is not substantially impacted by altering the parameters even up to an order of magnitude. This shows that our proposed framework is fairly robust to changes in input parameters which are inherently prone to misspecification.

**Table 10 Robustness checks on qol scores, costs, and weights for Events 1 and 2**

Policy: rolling-horizon	Monthly qol for experiencing Event 2 ( $q_{21}$ )	Monthly qol for experiencing Event 1 ( $q_{11}$ )						Comparison
		0.017 year		0.031 year		0.050 year		
		\$10K	\$100K	\$10K	\$100K	\$10K	\$100K	
(0.017, 0.020) <sup>†</sup>		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	vs. CDC 16
		1.00 (0.00)	0.97 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	vs. CDC 22
(0.033, 0.039)		1.00 (0.00)	0.93 (0.01)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	vs. CDC 16
		1.00 (0.00)	0.91 (0.02)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	vs. CDC 22
(0.050, 0.055)		1.00 (0.00)	0.88 (0.02)	1.00 (0.00)	0.93 (0.02)	1.00 (0.00)	0.99 (0.00)	vs. CDC 16
		1.00 (0.00)	0.86 (0.02)	1.00 (0.00)	0.88 (0.02)	1.00 (0.00)	0.94 (0.02)	vs. CDC 22

Policy: rolling-horizon	Cost of an occurrence of Event 2 ( $C_2$ )	Cost of an occurrence of Event 1 ( $C_1$ )						Comparison
		\$7,500		\$15,588.07		\$30,000		
		\$10K	\$100K	\$10K	\$100K	\$10K	\$100K	
(\$2,000, \$880) <sup>†</sup>		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.93 (0.02)	1.00 (0.00)	0.91 (0.02)	vs. CDC 16
		1.00 (0.00)	0.99 (0.01)	1.00 (0.00)	0.92 (0.01)	1.00 (0.00)	0.86 (0.01)	vs. CDC 22
(\$6,172.86, \$2,714.90)		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.94 (0.02)	vs. CDC 16
		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.91 (0.03)	vs. CDC 22
(\$20,000, \$8,800)		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	vs. CDC 16
		1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	0.93 (0.02)	vs. CDC 22

Policy: rolling-horizon	Weights for Events 1 and 2 ( $w_1, w_2$ )				Comparison	Weights for Events 1 and 2 ( $w_1, w_2$ )			
	\$10K	\$100K				\$10K	\$100K		
(1.00, 0.00)	0.39 (0.02)	0.29 (0.01)	vs. CDC 16		(0.25, 0.75)	1.00 (0.00)	1.00 (0.00)	vs. CDC 16	
	0.38 (0.02)	0.19 (0.01)	vs. CDC 22			1.00 (0.00)	1.00 (0.00)	vs. CDC 22	
(0.75, 0.25)	1.00 (0.00)	0.90 (0.02)	vs. CDC 16		(0.00, 1.00)	1.00 (0.00)	1.00 (0.00)	vs. CDC 16	
	1.00 (0.00)	0.67 (0.02)	vs. CDC 22			1.00 (0.00)	1.00 (0.00)	vs. CDC 22	
(0.50, 0.50)	1.00 (0.00)	1.00 (0.00)	vs. CDC 16						
	1.00 (0.00)	1.00 (0.00)	vs. CDC 22						

Notes. Results are the avg (s.d.) of CE probabilities from our rolling-horizon policy compared to the CDC 2016 and 2022 guidelines (with the dose-tapering rate  $\theta = 0.2$ ). Results from our myopic policy are presented in Appendix G. For qol scores, costs, and weights, 9, 9, and 5 scenarios are explored, respectively. Each scenario is evaluated under  $WTP \in \{\$10K, \$100K\}$ .  
<sup>†</sup>Numbers in ( ) represent the corresponding values for acute and chronic pain, respectively.

## 6. Policy Recommendations and Limitations

We now make use of our main findings and summarize a few important insights that could enable authorities impose more effective policies in addressing the opioid epidemic.

- The CDC has proposed two sets of guidelines in 2016 and 2022 addressing the opioid prescriptions for pain management. Our results show that the 2022 guidelines are more cost-effective than the 2016 ones (see §5.1.1). However, providing personalized algorithmic-based treatment plans by following an approach like the one we propose in this study can go a long way: it can drastically improve cost-effectiveness compared to both sets of CDC guidelines. As the CDC will soon evaluate the final set for the 2022 guidelines (NPR 2022), we hope that our results can inform authorities about adjustments needed to improve the suitability of their guidelines.
- Our results show that the guidelines should not underplay the importance of untreated or under-treated pain. For example, we find that, under the human-based approaches observed from the

practice, for each patient experiencing Event 1, there are on average about 24 patients experiencing Event 2. However, under our proposed treatment policies, in which the risks of Events 1 and 2 are optimized in a personal way, this number reduces to 10. Notably, we observe that focusing purely on reducing the instances of opioid misuse (the main objective of CDC guidelines) can be misleading. In particular, even if the guidelines were to hypothetically remove all instances of opioid misuse, our results indicate that the society would lose up to \$2,000 (\$1,300) worth of net monetary benefit for each patient suffering from acute (chronic) pain. This is because eliminating Event 1 would come at the expense of more frequent incidence of Event 2 (see §5.1.3). These findings shed light on the importance of using treatment policies that can make an effective balance (e.g., in a personalized manner) between the risks of both Events 1 and 2.

- In the absence of behavioral factors for patients (e.g., mental health disorder or substance abuse), the intensity of opioid dose could be higher than the amounts used in the practice. For this group, the opioid dose can be tapered down faster among younger female (compared to older male) patients or those with a history of surgery/inpatient admission pre-supply (compared to patients with no such history) (see §5.1.2).
- When patients have (do not have) a history of behavioral factors or surgery/inpatient admission pre-supply, the duration of medications supply should be higher for older male (younger female) patients than what we observe from the practice. Also, unlike the practice, the supply duration needs to increase in the early stage of therapeutic course and then decrease towards the end of this course. That is, given that the opioid dose is tapered down, the supply duration should be adjusted so as to offset the negative effects of the reduced dose (see §5.1.2).
- Compared to what we observe from the practice, non-pharmacologic treatments (NPHTs) should be prescribed at much higher rates. For acute pain treatment, NPHTs could be used for a few months after onset of pain, and any record of behavioral factors would elongate this period. For chronic pain treatment, NPHTs could be used over the whole (towards the end of) therapeutic course, if there is (is no) record of behavioral factors pre-supply (see §5.1.2).

Caution should be exercised when following our proposed algorithmic-based treatment policies. First, our framework outputs multi-modal pain treatments where both pharmacologic and non-pharmacologic options are taken into account. However, in the clinical practice, these treatments are typically administered by different providers, and some of such treatments are not covered by insurers. Therefore, further considerations should be taken to safeguard care coordination between different providers in charge as well as potential issues that might arise due to the lack of coverage by insurance companies. Second, to increase their adoption in a decision-support system, we first need to position analytics-driven algorithmic-based approaches in comparison with human-based approaches observed from the practice. While we accomplished this in our study, addressing issues such as the “degree of automation” (how much clinicians should be able to override our model-based recommendations?) or “human-in-the-loop” (how much the human judgment should

be combined with our algorithm-based approach?) is beyond the scope of the current study, and thus, we leave them for future research. Third, adopting our algorithmic-based approach requires various additional efforts which might be costly, including collecting and hosting large-scale training data sets, education providers, and ensuring adherence to the recommendations of the models. Despite all these, we believe the authorities should carefully consider adopting recommendations stemming from our framework as they may not only improve health outcomes (e.g., as measured by the QALYs gained), but also yield significant cost savings. For example, our results show that, while improving QALYs by 2.82 days per patient per year, the average cost saved per patient per year under our treatment policies compared to the observed practice of physicians is \$461.46 (see §5.1.1). Given that we obtain these results by averaging savings across over 3 million patients in our data set, we believe the overall benefits obtained from following an approach like ours is well worth the costs.

## 7. Conclusion

Instead of a human-based approach, in which assessing the risks and benefits of opioid use is left to individual physicians, we develop a comprehensive algorithmic-based approach, in which we employ longitudinal machine learning algorithms and train them on information pertaining more than 3 million patients. After testing these algorithms and identifying the best one, we feed it into a mathematical model that can help physicians in determining personalized and cost-effective multi-modal pain treatment plans.

We extensively compare the human-based approach as well as the guidelines provided by the CDC with our proposed algorithmic-based approach. Across various comparisons, our results show a superior performance for our proposed approach compared to both the CDC guidelines and the practice of physicians. Using findings from the comparisons between our algorithmic-based approach and the current human-based approach, we generate important insights for both physicians and policymakers.

Future research can advance our study via at least five potential avenues: (1) As more recent data becomes available, future research can better reflect on developments of other disease or pandemics (e.g., the COVID-19 pandemic) that might have (in)directly impacted the opioid crisis. (2) Due to the sheer volume of medical specialties and pain-inducing conditions (e.g., diseases of the musculoskeletal system and connective tissues), we do not break down our proposed treatment policies for each possible specialty. Future research can aid in using our framework and extending it separately for each specialty. (3) From a methodology standpoint, future research can improve our study by developing and applying other methods of dealing with the curse of dimensionality that arise when devising multi-period, multi-modal, and personalized pain treatment plans. (4) Full personalization of multi-period and multi-modal treatment plans via an algorithmic-based approach requires collecting various additional information from each patient (e.g., their genetic

profiles). Following an approach like the one we propose and training longitudinal machine learning models on such additional information can be yet another fruitful path for future research. (5) Finally, future research can explore better ways of combining human intuition with treatment plans we obtain from our algorithmic-based approach. This can yield an alternative method superior to both the human-based and the algorithmic-based approach we study in this work.

## References

- AMA (2016) AMA comments on CDC’s Proposed Opioids Prescription Policy. Retrieved Apr 10, 2021, [Link](#).
- Ataoglu E, Tiftik T, Kara M, Tunc H, Ersöz M, and others (2013) Effects of chronic pain on quality of life and depression in patients with spinal cord injury. *Spinal Cord*. 51(1):23–26.
- AZDHS: Arizona Department of Health Services (2017) Appendix B: State-by-state summary of opioid prescribing regulations and guidelines. Retrieved Dec 20, 2020, [Link](#).
- Barnett ML, Olenski AR, Jena AB (2017) Opioid-prescribing patterns of emergency physicians and risk of long-term use. *New England Journal of Medicine*. 376(7):663–673.
- Bastani H, Bayati M (2020) Online decision-making with high-dimensional covariates. *Operations Research*. 68(1):276–294.
- Bertsimas D, O’Hair A, Relyea S, Silberholz J (2016) An analytics approach to designing combination chemotherapy regimens for cancer. *Management Science*. 62(5):1511–1531.
- Bjarnadottir M, Anderson D, Agarwal R, Prasad K, Nelson A (2019) Continuity of care versus a second opinion: Evidence from the opioid crisis. *INFORMS Annual Conference*. Seattle, WA, USA. October 2019.
- Bolori A, Saghafian S, Chakkeri HA, Cook CB (2020a) Data-driven management of post-transplant medications: An ambiguous partially observable Markov decision process approach. *Manufacturing & Service Operations Management*. 22(5):1066–1087.
- Bolori A, Arnetz BB, Viens F, Maiti T, Arnetz JE (2020b) Misalignment of stakeholder incentives in the opioid crisis. *International Journal of Environmental Research and Public Health*. 17(20):7535.
- Capitaine L, Genuer R, Thiébaud R (2018) Random forests for high-dimensional longitudinal data. Retrieved Nov 10, 2020, [Link](#).
- Carr DB (2016) NPS versus CDC: Scylla, Charybdis and the “number needed to [under-] treat”. *Pain Medicine*. 17(6):999–1000.
- CDC: Centers for Disease Control and Prevention (2016) Pocket card for prescribing opioids for chronic pain. Retrieved Feb 16, 2020, [Link](#).
- CDC: Centers for Disease Control and Prevention (2020) Analyzing Opioid Prescription Data and Oral Morphine Milligram Equivalents (MME). Retrieved May 16, 2021, [Link](#).
- CDC: Centers for Disease Control and Prevention (2022) Federal Register Notice: CDC’s updated Clinical Practice Guideline for Prescribing Opioids is now open for public comment. Retrieved Apr 23, 2022, [Link](#).
- Che Z, Sauver JS, Liu H, Liu Y (2017) Deep learning solutions for classifying patients on opioid use. *AMIA Annual Symposium Proceedings*. 2017:525–534.
- Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J (2016) Doctor AI: Predicting clinical events via recurrent neural networks. *Machine Learning for Healthcare Conference*. 2016:301–318.
- Ciesielski T, Iyengar R, Bothra A, Tomala D, Cisló G, Gage BF (2016) A tool to assess risk of de novo opioid abuse or dependence. *The American Journal of Medicine*. 129(7):699–705.
- Cochran BN, Flentje A, Heck NC, Van Den Bos J, Perlman D and others (2014) Factors predicting development of opioid use disorders among individuals who receive an initial opioid prescription: mathematical modeling using a database of commercially-insured individuals. *Drug and Alcohol Dependence*. 138:202–208.
- Crosier BS, Borodovsky J, Mateu-Gelabert P, Guarino H (2017) Finding a needle in the haystack: Using machine-learning to predict overdose in opioid users. *Drug and Alcohol Dependence*. 171:e49.
- De Maeyer J, Vanderplasschen W, Broekaert E (2010) Quality of life among opiate-dependent individuals: A review of the literature. *International Journal of Drug Policy*. 21(5):364–380.
- Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*. 144(1):114.
- Domingo-Salvany A, Brugal MT, Barrio G, González-Saiz F, Bravo MJ, and others (2010) Gender differences in health related quality of life of young heroin users. *Health and Quality of Life Outcomes*. 8(1):1–10.
- Dowell D, Haegerich TM, Chou R (2016) CDC guideline for prescribing opioids for chronic pain—United States, 2016. *Jama*. 315(15):1624–1645.
- Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW (2015) *Methods for the economic evaluation of health care programmes*. Oxford University Press.
- FDA: U.S. Food and Drug Administration (2019) FDA identifies harm reported from sudden discontinuation of opioid pain medicines and requires label changes to guide prescribers on gradual, individualized tapering. Retrieved Nov 27, 2020, [Link](#).
- Fenwick E, Marshall DA, Levy AR, Nichol G (2006) Using and interpreting cost-effectiveness acceptability curves: an example using data from a trial of management strategies for atrial fibrillation. *BMC Health Services Research*. 6(1):52.
- Fishbain DA, Cole B, Lewis J, Rosomoff HL, Rosomoff RS (2007) What percentage of chronic nonmalignant pain patients exposed to chronic opioid analgesic therapy develop abuse/addiction and/or aberrant drug-related behaviors? A structured evidence-based review. *Pain Medicine*. 9(4):444–459.
- Freeman M, Bobroske K, Cattrell A, Huan L, Scholtes S (2019) Continuity of care versus a second opinion: Evidence from the opioid crisis. *M&SOM International Conference*. Singapore. July 2019.
- Gan K, Scheller-Wolf AA, Tayur SR (2019) Personalized Treatment for Opioid Use Disorder. Retrieved Jul 25, 2021, [Link](#).
- Giacomuzzi SM, Riemer Y, Ertl M, Kemmler G, Rössler H, and others (2005) Gender differences in health-related quality of life on admission to a maintenance treatment program. *European Addiction Research*. 11(2):69–75.
- Goertz CM, George SZ (2018) Insurer coverage of nonpharmacological treatments for low back pain—time for a change. *JAMA Network Open*. 1(6):e183037–e183037.



- Grover CA, Elder JW, Close RJH, Curry SM (2012) How frequently are “classic” drug-seeking behaviors used by drug-seeking patients in the emergency department? *Western Journal of Emergency Medicine*. 13(5):416–421.
- Hadi MA, McHugh GA, Closs SJ (2016) Impact of chronic pain on patients’ quality of life: a comparative mixed-methods study. *Journal of Patient Experience*. 6(2):133–141.
- Haller IV, Renier CM, Juusola M, Hitz P, Steffen W, and others (2016) Enhancing risk assessment in patients receiving chronic opioid analgesic therapy using natural language processing. *Pain Medicine*. 18(10):1952–1960.
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. *Springer*.
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Computation*. 9(8):1735–1780.
- Huskisson EC (1974) Measurement of pain. *The Lancet*. 304(7889):1127–1131.
- ICD9 (2015) ICD-9-CM Diagnosis Codes. Retrieved Jan 25, 2019, [Link](#).
- Johns Hopkins – The Bloomberg School of Public Health (2018) Health Insurer Policies May Discourage Use of Non-Opioid Alternatives for Lower Back Pain. Retrieved Aug 03, 2022, [Link](#).
- Johnson SP, Chung KC, Zhong L, Shauver MJ, Engelsbe MJ, and others (2016) Risk of prolonged opioid use among opioid-naïve patients following common hand surgery procedures. *The Journal of Hand Surgery*. 41(10):947–957.
- Katz J, Melzack R (1999) Measurement of pain. *Surgical Clinics of North America*. 79(2):231–252.
- Katz N (2002) The impact of pain management on quality of life. *Journal of Pain and Symptom Management*. 24(1):S38–S47.
- Keane PA, Topol EJ (2021) AI-facilitated health care requires education of clinicians. *The Lancet*. 397(10281):1254.
- Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D (2019) Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*. 17(1):1–9.
- Luo F, Li M, Florence C (2021) State-level economic costs of opioid use disorder and fatal opioid overdose—United States, 2017. *Morbidity and Mortality Weekly Report*. 70(15):541–546.
- Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*. 151:90–103.
- McPhillips-Tangum CA, Cherkin DC, Rhodes LA, Markham C (1998) Reasons for repeated medical visits among patients with chronic back pain. Announcement. *Journal of General Internal Medicine*. 13(5):289–295.
- Mišić VV (2020) Optimization of tree ensembles. *Operations Research*. 68(5):1605–1624.
- Mišić VV, Perakis G (2020) Data analytics in operations management: A review. *Manufacturing & Service Operations Management*. 22(1):158–169.
- NIDA: National Institute on Drug Abuse (2022) Overdose Death Rates. Retrieved Jul 07, 2022, [Link](#).
- NPR: National Public Radio (2022) CDC weighs new opioid prescribing guidelines amid controversy over old ones. Retrieved Apr 21, 2022, [Link](#).
- Nuckols TK, Anderson L, Popescu I, Diamant AL, Doyle B, and others (2014) Opioid prescribing: a systematic review and critical appraisal of guidelines for chronic pain. *Annals of Internal Medicine*. 160(1):38–47.
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature*. 323:533–536.
- Palliative Drugs (2009) Opioid conversions. Retrieved Jan 10, 2019, [Link](#).
- Pitt AL, Humphreys K, Brandeau ML (2018) Modeling health benefits and harms of public policy responses to the US opioid epidemic. *American Journal of Public Health*. 108(10):1394–1400.
- Pletcher MJ, Kertesz SG, Kohn MA, Gonzales R (2008) Trends in opioid prescribing by race/ethnicity for patients seeking care in US emergency departments. *JAMA*. 299(1):70–78.
- Saghafian S (2018) Ambiguous partially observable Markov decision processes: Structural results and applications. *Journal of Economic Theory*. 178:1–35.
- Saghafian S, Murphy SA (2021) Innovative health care delivery: The scientific and regulatory challenges in designing mHealth interventions. *NAM Perspectives*.
- Saghafian S (2022) Ambiguous Dynamic Treatment Regimes: A Reinforcement Learning Approach. Retrieved Sep 06, 2022, [Link](#).
- Sexton J (2018) Historical Tree Ensembles for Longitudinal Data. Retrieved Mar 15, 2020, [Link](#).
- Smith CA, Levett KM, Collins CT, Armour M, Dahlen HG, Sukanuma M (2018) Relaxation techniques for pain management in labour. *Cochrane Database of Systematic Reviews*. Issue 3. Art. No.: CD009514. DOI: 10.1002/14651858.CD009514.pub2.
- Taylor RS, Ullrich K, Regan S, Broussard C, Schwenkglenks M, and others (2013) The impact of early postoperative pain on health-related quality of life. *Pain Practice*. 13(7):515–523.
- Teater D (2015) Evidence for the efficacy of pain medications. Retrieved Dec 10, 2019, [Link](#).
- Tüzün E (2007) Quality of life in chronic musculoskeletal pain. *Best Practice & Research Clinical Rheumatology*. 21(3):567–579.
- USA Today (2016) Doctors told to avoid prescribing opiates for chronic pain. Retrieved Sep 01, 2020, [Link](#).
- USIC: US Inflation Calculator (2021) Health Care Inflation in the United States (1948-2021). Retrieved Nov 20, 2021, [Link](#).
- VA.gov: U.S. Department of Veteran Affairs (2016) Pain Management Opioid Taper Decision Tool. Retrieved Nov 10, 2021, [Link](#).
- Vunikili R, Glicksberg BS, Johnson KW, Dudley J, Subramanian L, Khader S (2018) Predictive modeling of susceptibility to substance abuse, mortality and drug-drug interactions in opioid patients. Retrieved Mar 01, 2020, [Link](#).
- Wells N, Pasero C, McCaffery M (2008) Improving the quality of care through pain assessment and management. In *Patient safety and quality: An evidence-based handbook for nurses*. Agency for Healthcare Research and Quality (US).
- Wilson JR, Lorenz KA (2015) Modeling Binary Correlated Responses using SAS, SPSS and R. *Springer*. 103–130.
- Wu CL, Naqibuddin M, Rowlingson AJ, Lietman SA, Jermyn RM, and others (2003) The effect of pain on health-related quality of life in the immediate postoperative period. *Anesthesia & Analgesia*. 97(4):1078–1085.
- Wu CL, King AB, Geiger TM, Grant MC, Grocott MP, and others (2019) American society for enhanced recovery and perioperative quality initiative joint consensus statement on perioperative opioid minimization in opioid-naïve patients. *Anesthesia and Analgesia*. 129(2):567.
- Xiao H, Barber JP (2008) The effect of perceived health status on patient satisfaction. *Value in Health*. 11(4):719–725.