Modern Methods in Precision Medicine

Kyra (Jingyi) Gan

April 2022

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:

Andrew Li (Co-chair)
Sridhar Tayur (Co-chair)
Tinglong Dai
Zachary Lipton
Alan Scheller-Wolf

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Operations Research.

Contents

Chapter 1

Introduction

As the concept of precision medicine spreads, there is a growing need for developing better algorithms that a) are sample e cient (i.e., require fewer samples to achieve the same accuracy level), b) think beyond association (to identify the causation hidden in the data), and c) provide insights to medical practice. In this dissertation, we investigate various problems in precision medicine, the topics ranging from opioid use disorder (OUD) and cancer treatment, to sickle cell disease (SCD). We leverage tools from stochastic learning, causal inference, and machine learning, with the objective of reducing healthcare expenditure and improving the quality of care.

One of the US's most recent health crises is the opioid overdose epidemic, and the resource that we have to reverse is epidemic is limited. While various OUD treatments have shown to be e ective on a population level, individual patients react di erently to these treatments. Wearable devices, on the other hand, can potentially revolutionize treatments for OUD by measuring patient responses to di erent treatment regimens in real-time, enabling the development of personalized treatments. However, before we deploy the use of wearable devices in OUD treatments, we rst need to understand the practicality and the cost-e ectiveness of such devices. Thus motivated, in Chapter 2, we evaluate the use of wearable devices in OUD treatments when the budget is limited. In particular, we consider a variety of wearable devices with di erent features, sensitivities, and costs, and model our problem using a nite-horizon, non-stationary, constrained partially observable Markov decision process (CPOMDP). To facilitate the solution of our model, we provide a novel budget reformulation that nds all optimal solutions lying on the original formulation's solution's convex hull. Next, we show our reformulation can be solved using a binary search in conjunction with an exact POMDP algorithm. We apply those elements, using extracted transition matrices and rewards from past literature, to perform a numerical study to investigate the value of incorporating di erent wearables in treatments for OUD under scenarios described by di erent levels of budget, wearable precision, and patient treatment adherence (TA). We nd that wearables can be valuable at moderate budgets for patients with low or moderate TA, this bene t increases

as the wearable accuracy increases. Outside of these settings, either the marginal bene t of wearables is negligible relative to their cost, or their use increases the patients' risk of overdose to an unacceptable degree.

Chapters 3 and 4 both relate to cancer research. One of the fundamental goals in cancer research is to identify the genetic mutations that cacausecancer. If such mutations were identi ed, then targeted drugs can be produced to block the e ect of these mutations and hence curing cancer. Since editing human genome is clinically unsafe at the currently stage, to derive such causal relations, we can only useobservationaldata collected from a patient population of interest. Motivated by the fact that the majority of patients only have a subset of genes sequenced,[1] in Chapter 3, we consider the bene t of incorporating a largeconfoundedobservational dataset (onfounder unobserved) alongside a smalldeconfoundedobservational dataset (onfounder revealed) when estimating the average treatment e ect(ATE). Our theoretical results show that the inclusion of confounded data can signi cantly reduce the quantity of deconfounded data required to estimate the ATE to within a desired accuracy level. Moreover, in some cases say, genetics we could imagine retrospectively selecting samples to deconfound. We demonstrate that by actively selecting these samples based upon the (already observed) treatment and outcome, we can reduce our data dependence further. Our theoretical results establish that the worst-case relative performance of our approach (vs. random selection) is bounded while our best-case gains are unbounded. We perform extensive synthetic experiments to validate our theoretical results. Finally, we demonstrate the practical bene ts of selective deconfounding using a large real-world dataset related to genetic mutation in cancer.

Chapter 4 focuses on liquid biopsies simple blood tests that can be used for accurate early stage cancer detection. In particular, we study a set of problems that occur in the development of liquid biopsies via the lens oactive sequential hypothesis testing(ASHT). In the problem of ASHT, a learner seeks to identify thetrue hypothesis from among a known set of hypotheses. The learner is given a set of actions and knows the random distribution of the outcome of any action under any true hypothesis. Given a target error> 0, the goal is to sequentially select the fewest number of actions so as to identify the true hypothesis with probability at leas1t . Motivated by applications in which the number of hypotheses or actions is massive (e.g., genomics-based cancer detection), we propose e cient (greedy, in fact) algorithms and provide the rst approximation guarantees for ASHT, under two types of adaptivity. Both of our guarantees are independent of the number of actions and logarithmic in the number of hypotheses. We numerically evaluate the performance of our algorithms using both synthetic and real-world DNA mutation data, demonstrating that our algorithms outperform previously proposed heuristic policies by large

[1] Often this subset is the same across the patient population because doctors will only order a gene to be sequenced if there are known treatments for that gene.

margins.

Finally, Chapter 5 is an empirical chapter, where we focus on solving real-world problems where we collaborate with physicians. This chapter is motivated by improving the gap between machine learning research in healthcare and what has been implemented in practice. In particular, we collaborated closely with Dr. Patel and Dr. Novelli from University of Pittsburgh Medical Center in predicting the 30-day readmission risk for patients with sickle cell disease. Reducing preventable hospital readmissions in SCD could potentially improve outcomes and decrease healthcare costs. In a retrospective study of electronic health records, we hypothesized machine learning (ML) algorithms may outperform standard readmission scoring systems (LACE and HOSPITAL indices). Participants (n=446) included patients with SCD with at least one unplanned inpatient encounter between January 1, 2013, and November 1, 2018. Patients were randomly partitioned into training and testing groups. Unplanned hospital admissions (n=3299) were strati ed to training and testing samples. Potential predictors (n=486), measured from the last unplanned inpatient discharge to the current unplanned inpatient visit, were obtained via both data-driven methods and clinical knowledge. Three standard ML algorithms logistic regression (LR), support vector machine (SVM), and random forest (RF) were applied. Prediction performance was assessed using the C-statistic, sensitivity, and speci city. In this dataset, ML algorithms outperformed LACE (C-statistic 0.6, 95%CI 0.57-0.64) and HOSPITAL (C-statistic 0.69, 95%CI 0.66-0.72), with the RF (C-statistic 0.77, 95%CI 0.73-0.79) and LR (C-statistic 0.77, 95%CI 0.73-0.8) performing the best. We reported the most important predictors in our best models, and derive clinical insights.

1.1 Review of Classical Results in Concentration Inequalities for Subgaussian Random Variables

We rst state some classic results that we will use frequently in Chapters 3 and 4. We will consider the commonly used subgaussian distributions (Vershynin 2018). Loosely speaking, a random variable is subgaussian if its tail vanishes at a rate faster than some Gaussian distributions.

De nition 1 (subgaussian norm) Let X be a random variable, its subgaussian norm is de ned as $\|X\|_{\psi_2} := \inf\{t : E[e^{X^2/t^2}] \le 2\}$. Moreover, X is called subgaussian if $\|X\|_{\psi_2} < \infty$.

Many commonly used distributions satisfy this assumption, e.g., Bernoulli, uniform, and Gaussian distributions etc. We introduce a standard concentration bound for subgaussian random variables.

Theorem 1 (Hoe ding Inequality Vershynin 2018) Let $X_1, \ldots, X_n$ be independent subgaussian

random variables. Then for any $\gamma > 0$, it holds that

$$P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mathbb{E}X_i\right| \geq \gamma\right) \leq 2\exp\left(-\frac{2\gamma^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

1.1.1 Special Case: Bounded Random Variables

In addition to being subgaussian, if the random variable is bounded, then a stronger version of Theorem 1 can be stated as follows:

Lemma 1 (Hoeffding's Lemma) Let X be any real-valued random variable with expected value $\mathbb{E}[X] = 0$, such that $a \leq X \leq b$ almost surely. Then, for all $\lambda \in \mathbb{R}$, $\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2 (b-a)^2}{8}\right)$.

Theorem 2 (Hoeffding's inequality for general bounded r.v.s) Let $X_1, \ldots, X_N$ be independent random variables such that $X_i \in [m_i, M_i]$, $\forall i$. Then, for $t > 0$, we have $P\left(\left|\sum_{i=1}^{N} (X_i - \mathbb{E}[X_i])\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right)$.

Chapter 2

Personalized Treatment for Opioid Use Disorder

2.1 Introduction

The national opioid use disorder crisis in the United States leads to thousands of death annually (Skolnick 2018), a ecting populations from all demographics. Since repeated opioid use can alter how we perceive motivation and reward long-term (Humphreys et al. 2017), making a full recovery from OUD is di cult and typically costly. In the United States, medication assisted treatment (MAT) the standard treatment for OUD includes the use of medications in combination with counseling and behavioral therapies. E ective treatment is made di cult by the fact that patients react to medications for OUD di erently, for example, due to genetic variations (Crist et al. 2013), cultural or ethnic di erences (Campbell and Edwards 2012), and stress (Sinha 2008). As a result, the treatment retention rate among patients with OUD remains low while the death rate remains high: studies have shown that the 2-month retention rate among patients with OUD is 57% with few staying enrolled beyond 3 months (Skolnick 2018), and the 30-year (after the initialization of the rst treatments) death rate is 47% (Grella and Lovinger 2011). Thus, there is a need to develop better, personalized, treatment for OUD.

One key step in developing personalized treatment for OUD is to measure patients' treatment responses. While most outpatient programs estimate the e ectiveness of a treatment through relapse rate (via urine tests), ecological momentary assessment (EMA) studies in which patients respond to daily surveys nd that addressing craving episodes before relapse can likely help prevent actual relapse (Serre et al. (2012, 2015), Epstein et al. (2009)). However, EMA is not reliable in detecting cravings because it is subject to response bias: for example, adolescents tend to provide random information (McLellan et al. 1992), and patients ashamed about cravings may

provide falsi ed information (Kleber et al. 2006).

In this work, we provide a framework to investigate the bene t of incorporating wearable devices in treatments for OUD, where wearable devices are de ned as smart electronic devices that can be worn on wrists to collect data. There is an emerging trend of integrating wearable devices in medical treatments (Cheol Jeong et al. 2018), and pilot studies demonstrate that wearables can be useful in treating Parkinson's disease (Suzuki et al. 2017), post-discharge monitoring of ICU patients (Kroll et al. 2017), and detecting early-stage Alzheimer's disease (Varatharajan et al. 2019). Moreover, Fatseas et al. (2011, 2015) show that relapses and strong cravings could potentially be captured by the sensors contained inside the wearables. As a result, many start-ups (Valant et al. 2018, Linder 2019) are integrating wearables into treatments for OUD, and there is thus an urgent need to develop tools for assessing the value of wearable devices in those treatments. Throughout this work, we assume there exists an algorithm that detects patient health states in real-time, possibly with some uncertainty, using the features captured by wearables.

However, the potential advantage of wearables is constrained by the limited national budget to ght this epidemic (NIDA 2020). Speci cally, it is unclear whether reducing the amount of money spent on MAT in favor of buying wearables is cost-e ective. Complicating this problem, there is a variety of wearables with di erent prices, sensors, and accuracies available. In this work, we provide a framework to assess the cost-bene t trade-o of di erent wearables from the perspective of the healthcare system, to help determine whether and which wearables should be invested in by treatment programs for OUD.

Our contributions in this research are threefold. First, we provide a framework for understanding the values of di erent wearable devices in OUD treatments under budget constraints. Since patient health states are not always fully observable, we formulate a budget-constrained, discrete-time, nite-horizon, non-stationary partially observable Markov decision process. We consider three classes of MAT treatments in addition to counseling only and no treatment, incorporating di erent transition matrices to model cases in which we have no wearables, have wearables that provide di erent levels of information accuracy on patient health states, or have wearables that provide perfect information, i.e., a full information MDP benchmark model.

Second, we provide a novel budget reformulation for our CPOMDP that could potentially be applied to other CPOMDP models. To our knowledge, only two works have proposed methods to solve nite-horizon CPOMDPs (Cevik et al. 2018, Undurti and How 2010). However, the former has no feasibility guarantees, and neither have optimality guarantees. In contrast, our budget reformulation nds all optimal solutions lying on the convex hull of the original formulation's solutions. Moreover, our reformulation can be solved using a binary search in conjunction with an exact POMDP algorithm. (Similarly, in the case of CMDP, our formulation can be solved using a binary search in conjunction with an exact MDP algorithm.) We show that our reformulation

not only guarantees the feasibility of our solution, but also optimality when randomized policies are allowed.

Third, we conduct a numerical study from the perspective of the healthcare system to provide insights that could potentially guide future eld studies. Incorporating di erent device costs and accuracies, our objective is to maximize the total lifetime discounted quality-adjusted life days(QALDs) of patients subject to the budget constraint. We discover that the health bene t of incorporating wearables could be signi cant when the budget is not very generous, because if the budget is very generous, we can a ord the most expensive (and e ective) treatment in every period, and the marginal bene t of wearables is negligible. Furthermore, assuming that wearables do not a ect patient treatment adherence levels, di erent patient types obtain di erent levels of bene t from wearables: patients with the highest treatment adherence bene t the least from wearable devices at all budget levels, and patients with lowTAs bene t the most when the budget is relatively low.

The paper is organized as follows. After a literature review in Ÿ 2.2, we formulate our model in Ÿ 2.4. In Ÿ 2.5 we introduce our budget reformulation, review an exact solution method for solving unconstrained POMDPs, and provide algorithms to solve our reformulation. In addition, we develop and introduce a heuristic algorithm to speed up the solving time of our reformulation, and provide a worst case error bound in Ÿ 2.15. In Ÿ 2.6 we compare wearables with various features and accuracies, and discuss our numerical results. We conclude in Ÿ 2.7. A notation table is included in Ÿ 2.3.

2.2 Literature Review

OUD Treatments In addition to the reasons listed in the introduction (genetics, cultural and ethnic di erences, and stress), the e ectiveness of a treatment for OUD can be in uenced by comorbid medical conditions (Luo and Levin 2017), age (McLellan et al. 1994), co-occurring mental health disorders (Morse and Bride 2017), education levels, psychiatric functioning, marital separation or social functioning (Sayre et al. 2002), treatment enrollment duration (Eastwood et al. 2017), and multi-drug usage (Williamson et al. 2006). The breath of these features highlights the potential of personalized treatments.

Many tools have been investigated to try to reverse the opioid epidemic. These include strengthening the regulations for opioid prescription (Kolodny and Frieden 2017), predicting opioid overdose via machine learning (Lo-Ciganic et al. 2019), and increasing the accessibility of OUD treatments (Marshall et al. 2015). However, none address e cient treatments. Several studies established that some treatment for OUD is more cost-e ective than no treatment, due to reduced hospital visits (e.g. Baser et al. (2011)), using either statistical tools or simulation models.

None explored the use of personalized treatment.

MDP and POMDP in Personalized Treatment A few works in the Operations Research/-Management Science literature have considered budget constraints in treatment decision models. Ayvaci et al. (2012) and Cevik et al. (2018) used nite horizon MDP and POMDP respectively to model the optimal breast cancer screening policy under budget constraints. Chen et al. (2018) formulated a nite horizon POMDP to model optimal liver cancer screening policies for patients with hepatitis C infection under the constraint that the policies can change at most a given number of times, and conducted numerical experiments using a MDP. All three papers reformulated their problems into mixed integer linear programs (MILPs). Ayvaci et al. (2012) showed that the optimal patient health outcome is strictly concave with respect to budget if randomized policies were allowed; to enforce a deterministic optimal solution, they must add integrality constraints in their MILP model. Since the space of reachable belief states in a POMDP model is in nite, the reformulation proposed by Cevik et al. (2018) is computationally intractable. As a result, they discretized their belief space and obtained an approximate solution. Chen et al. (2018) showed that the marginal bene t of surveillance is higher in patient populations with faster disease progressions. Furthermore, as patients' risk of developing cancer diminishes and their health outcome improves, the frequency of surveillance should not increase.

Additional studies have used either an MDP or a POMDP to model clinical decisions, with the majority focusing on maximizing quality-adjusted life years (QALYs). Zhang et al. (2012) studied optimal prostate biopsy referral decisions using a in nite-horizon, non-stationary POMDP. Ayer et al. (2012, 2015) and Alagoz et al. (2013) studied the optimal clinical decisions related to breast cancer using nite-horizon POMDPs, and MDPs respectively. Erenay et al. (2014) and Suen et al. (2017) studied optimal colonoscopy screening and optimal drug sensitivity test in tuberculosis treatment, respectively, using nite-horizon, non-stationary POMDPs.

CPOMDPs in the Computer Science Literature Incorporating a cost constraint directly into a POMDP yields a model that is computational intractable. Two lines of work in the computer science literature address this problem in the setting of in nite horizon CPOMDPs and nite horizon CPOMDPs, respectively, with the majority focusing on the former. Both directions heavily rely on reformulating the problem into either an MILP or a linear program (LP). Within the in nite horizon setting, Isom et al. (2008) modi ed the pruning step in an exact algorithm for solving unconstrained POMDPs to incorporate the constraint. Kim et al. (2011) proposed a heuristic that uses point-based value iteration in conjunction with an LP. Poupart et al. (2015) converted the problem into an LP and considered only a subset of belief states. To ensure optimality, they proposed an iterative algorithm to enlarge the subset of the belief states. However, this method cannot be adopted into

the nite-horizon setting. Lee et al. (2018) assumed unknown transition matrices and proposed to solve an unconstrained POMDP while optimizing its LP-induced parameters that control the trade-o between rewards and costs. None of the methods above have feasibility guarantees, and the majority do not guarantee that the nal solution can be made su ciently close to the optimal solution. In the nite horizon setting, Undurti and How (2010) proposed an algorithm that combines an o ine nite lookahead and an online branch-and-bound algorithm to ensure the feasibility of the solution, however without optimality guarantees. Furthermore, to calculate the expected reward and cost, all methods mentioned above must either solve a system of linear programs or conduct simulations.

In contrast, our nite-horizon budget reformulation guarantees the feasibility of the nal solution and admits optimal solutions when randomized policies are allowed; when deterministic solutions must be enforced, our reformulation might nd a sub-optimal solution. In addition, we show that our reformulation can be solved using a binary search in conjunction with an exact POMDP (or MDP) algorithm.

2.3 Notation Table

Notation	Description
TA	treatment adherence
MAT	medication assisted treatment
EMA	ecological momentary assessment
QALD	quality-adjusted life days
QALY	quality-adjusted life years
M	methadone maintenance treatments with counseling
B	buprenorphine maintenance treatments with counseling
NT; IN; CO	no treatment; implant naltrexone with counseling; counseling only
Re, Dx, OD, Dt	states of replase, detoxi cation, overdose, and death respectively
NC, C1, C2	states of no craving, low craving, and high craving intensity, respectively
Abs	absorbing state
ATD	average treatment dynamics
PTD	personalized treatment dynamics
UT	urine test
N	number of treatment periods
S	the set of all information states
A	the set of all actions

O	the set of all observations
$P(s_{t+1} \mid s_t, a)$	transition probability under action a at time t
P_t^a	transition probability matrix under action a at time t
$w(o_t \mid s_t, a)$; W	observation probability; matrix containing $w(o_t \mid s_t, a)$
;	sensitivity ; 1* sensitivity
;	speci city ; 1* speci city
; B	belief state; the set of all belief states
$h_t^a(s)$; $c_t^a(s)$	health reward and cost at state s under action a at time t, respectively
$r_t(a, s)$	immediate reward at state s under action a at time t
$_t$	amount of budget allocated for the rest of horizon in month t
$V_t^<(\ _t)$	optimal expected value of the objective function at time t under belief $_t$
$P(o_{t+1} \mid a, \ _t)$	probability of observing o_{t+1} after taking action a at belief state $_t$
	set of all feasible policies
$^{\cdot}(\ _t)$	optimal treatment policy under budget $_t$ (in System (1))
$_0$	initial belief state
$H_{\ _t}$, $C_{\ _t}$	expected health and cost under belief $_t$ and policy , respectively
	tunable parameter that takes values between 0 and 1
$V(\ _t;\ )$	expected reward under policy , belief $_t$, and
$^<$	optimal policy in System (2)
$^<$	optimal in System (2)
$^<$	optimal solution set to the unconstrained POMDP in System (2) under
$^<$	solution with the lowest expected health in $^<$
$V^<(\ _0;\ _t)$	optimal value in System (2) under budget $_t$
$^<_<$	optimal policy with the lowest expected health under $^<$ in Equation (2.9)
$F(\ _t)$	obtained by connecting the end points of the step function $H_{^{\cdot}(\ _0)}(\ _t)$
$F(\ _t)$	point-wise smallest concave function whose hypograph contains that of $(\ _t)$
$_t$	vector containing the expected reward at each state s_t
$A_t^{a,o}$	minimal representation of the -vector set for action a_t, observation o_{t+1}
A_t^a	minimal representation of $_{o_{t+1} \in O_{t+1}} A_t^{a,o}$
A_t	minimal representation of $_{a_t \in A_t} A_t^a$
$r(\ _{t+1}, a_t, o_{t+1}, s_t)$	scaled expected reward at state s_t given $_{t+1}$ after taking a_t and observing o_{t+1}
$_t$	the vector containing the expected cost at each state s_t
x_s; nc_s; $c2_s$;	probability of transitioning to Dx, NC, C2 from state s, respectively
e_s; od_s; d_s	probability of transitioning to Re, OD, Dt from state s, respectively
w	probability of withdrawing
$\hat{T}_0^{a'}$	transition matrices for ATD

2.4 Constrained Partially Observable Markov Decision Process

In this section we describe our CPOMDP model, where a CPOMDP is de ned as a POMDP with two additional components (Isom et al. 2008): 1) a cost incurred in each state for executing an action, and 2) an upper bound on the cumulative cost. We consider three di erent cases: (1) without any wearables; (2) with wearable devices that can detect cravings with various levels of accuracy; and (3) with wearable devices that can capture craving episodes with perfect accuracy. For each case, the objective of our model is to maximize the QALDs for an individual patient subject to a prede ned budget constraint for the patient (which would be a fraction of the overall budget).

Time Horizon Let N denote the number of treatment periods. Although our model can be solved for longer horizons, to keep our model parsimonious, we use a one-year horizon to mimic the federal budget allocation for treating OUD. We further discretize the horizon into twelve months ($N = 12$) the recommended change in treatments by the American Psychiatric Association (Kleber et al. 2006) is less than once per month. Within each month, a patient can transition between di erent states. Because the natural granularity of the clinical data makes daily treatment transition parameters easier to de ne and estimate than monthly ones, our transition probabilities and immediate rewards are de ned in days.

Actions According to the American Society of Addiction Medicine (ASAM 2016), methadone, buprenorphine, and naltrexone are three standard medical treatments for OUD; these treatments are typically provided along with counseling and other support. Therefore, at the beginning of each month, a care provider can decide which one of the following ve actions to take: no treatment (NT), methadone maintenance treatment with counseling (M), buprenorphine maintenance treatment with counseling (B), implant naltrexone with counseling (IN), or counseling only (CO). Because treatments M and B include prescribing medicine on a daily basis, their treatment outcomes are positively correlated with patient treatment adherence levels. Treatment IN requires only monthly implant procedures, and thus it works the same for all TA groups. Several treatment constraints mentioned by Kleber et al. (2006) are not implemented in our model but can be added easily: for example, treatment B is not suitable for patients with liver disease.

States The information states $S(t)$ in our models correspond to patient health states. We do not de ne the state recovery in our model because OUD is unlikely to be cured within one year: from our conversion with practitioners, a patient faces the risk of relapse even after 10 years of abstinence. We assume that if a patient starts to use drugs again (relapses, Re) in the program, we will allow him or her to stay in the program. If a patient relapses, he or she can either go through a detoxi cation (Dx) program to stop using drugs or stop on his or her own (Zarkin et al. 2005). Because our optimization problem terminates once a patient withdraws from the program, we create an absorbing state (Abs) representing that the patient has either withdrawn from the program or died. In any given in-treatment day, (i.e., a day outside of state Dt, Dx, or Abs,) a patient can experience either no craving (NC) for drugs the entire day or some craving at certain points of the day; clinical papers have found that not all cravings lead to relapse (Marsden et al. 2014, Serre et al. 2018). Based on conversations with practitioners, we denote a low craving intensity state (C1) and a high craving intensity state (C2); the patient is more likely to relapse in state C2 than in C1, and in state C1 than in NC. Any renewed opioid usage after a period of abstinence carries an increased risk of overdose (OD) requiring medical attention due to loss of tolerance (Chalana et al. 2016), which can lead to death (Dt). Thus, the health of a patient falls into one of the following eight states: $S := \{Dx, NC, C1, C2, Re, OD, Dt, Abs\} := S$.

Transition Probability Matrices The transition probability $P(s_{t+1}|s_t, a)$ is the transition probability from state $s_t \in S$ to state $s_{t+1} \in S$ under action $a \in A := \{NT, M, B, IN, CO\}$, where t indicates the number of days since the last (known) drug-use.[1] To re ect di erent wearable accuracies in detecting cravings and estimating individual reactions to treatments, we perturb the observation and transition matrices governing our POMDP, respectively. We describe how the transition probabilities were estimated or generated in § 2.6.2.

Observations and Observation Matrices Let $o_t \in O$ denote the observation at time t. In our model, the set of feasible observations, $O := \{Dx, NC, C1, C2, Re, OD, Dt, Abs\}$[2], is the same for all actions at all t. At every period, we perform urine test to decide whether a patient has relapsed or not. Since this assumption is potentially more conservative than necessary (as we could reduce the frequency of urine tests if the wearable is su ciently accurate), we relax this assumption in § 2.6.4. We assume that urine tests can accurately detect drug usage within a three-day interval (Lautieri 2019). Let $w(o_t|s_t, a)$ be the probability of making observation o_t in state s_t under action a, and

[1] To keep the our model Markovian, the transition probability out of the state detoxi cation was modeled as a geometric distribution as indicated by Table 2.6.

[2] Note that as we will see in § 2.6, it is important that O is the same as S for the consistency of the model evaluation.

let W denote the matrix containing $w.o_t|s_t; a/$, with columns corresponding to observations and rows to true states.

Case 1 no wearables: When there is no wearable, we only observe either a negative urine test result, ut* , or a positive result, ut+. Since we do not observe patients' craving states states NC, C1, and C2 we maintain a uniform belief over these three states if we observe a negative urine test. However, a positive urine test result does not necessarily indicate that the patient is in the state relapse because a patient can stop using drugs on his or her own. Thus, a care provider can partially observe state Re and has no information about states NC, C1, and C2 in this model. Let nc_{ut+}, $C1_{ut+}$, and $C2_{ut+}$ denote the probability of observing a positive urine test when the patient is in fact in states NC, C1, and C2, respectively. Then,

$$W = \begin{array}{c} \\ \text{Dx} \\ \text{NC} \\ \text{C1} \\ \text{C2} \\ \text{Re} \\ \text{OD} \\ \text{Dt} \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Dx} & \text{NC} & \text{C1} & \text{C2} & \text{Re} & \text{OD} & \text{Dt} & \text{Abs} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .1 * nc_{ut+}/\_3 & .1 * nc_{ut+}/\_3 & .1 * nc_{ut+}/\_3 & nc_{ut+} & 0 & 0 & 0 \\ 0 & .1 * C1_{ut+}/\_3 & .1 * C1_{ut+}/\_3 & .1 * C1_{ut+}/\_3 & C1_{ut+} & 0 & 0 & 0 \\ 0 & .1 * C2_{ut+}/\_3 & .1 * C2_{ut+}/\_3 & .1 * C2_{ut+}/\_3 & C2_{ut+} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

The calculations of nc_{ut+} and $C1_{ut+}$ are included in Ÿ 2.8. After plugging in our estimated transition matrices derived from past literature, $nc_{ut+}; C1_{ut+}; C2_{ut+} < 0:01$ since the probability that a patient recovers from a relapse within 3 days is very small (Zarkin et al. 2005). Therefore, the state Re is fully observable in this case.

Case 2 wearables with imperfect information: In this case, we assume there exists an algorithm that takes urine test results and data collected via the wearables as inputs and returns an estimate of the patient's current health state. However, depending on the accuracy and dimension of the inputs, the algorithm will have di erent sensitivities and speci cities for each partially observable state, where the sensitivity ($_s$) is the probability of observing states Ë S given that the patient is in state s, and the speci city (p_s) is the probability of not observing state s given that the patient is not in state s. To simplify the representation of W, we parametrize[3] W as follows, where we put more weights on worse health states:

[3] There are many equivalent parameterization of this problem.

$$W = \begin{array}{c} s\_o \\ Dx \\ NC \\ C1 \\ C2 \\ Re \\ OD \\ Dt \\ Abs \end{array} \begin{array}{c} \begin{array}{cccccccc} Dx & NC & C1 & C2 & Re & OD & Dt & Abs \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{Re}p_{C2}p_{C1} & p_{Re}p_{C2}\neg p_{C1} & p_{Re}\neg p_{C2}p_{C1} & \neg p_{Re} & 0 & 0 & 0 \\ 0 & p_{Re}p_{C2}\neg p_{C1} & p_{Re}p_{C2}\ C1 & p_{Re}\neg p_{C2}\ C1 & \neg p_{Re} & 0 & 0 & 0 \\ 0 & p_{Re}\neg p_{C2}p_{C1} & p_{Re}\neg p_{C2}\neg p_{C1} & p_{Re}\ C2 & \neg p_{Re} & 0 & 0 & 0 \\ 0 & \neg p_{Re}p_{C2}p_{C1} & \neg p_{Re}p_{C2}\neg p_{C1} & \neg p_{Re}\neg p_{C2} & Re & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

where $\neg p_s = 1 * {}_s$, and $\neg p_s = 1 * p_s$, for $s \in \{Re, C1, C2\}$. As in the previous case, Re is fully observed if monthly urine tests are present: we set $p_{Re} = {}_{Re} = 1$.

Case 3 wearables with perfect information: In this case, wearables can correctly detect a patient's health state. Thus, the observation matrix is the identity matrix, and the model becomes a Markov decision process.

Belief States The belief state at time t, $_t = (_t(Dx), _t(NC), _t(C1), _t(C2), _t(Re), _t(OD), _t(Dt), _t(Abs)) \in B$, defines the probabilities that the care provider believes the patient is in before any action is taken. If the patient is in a fully observable state s, then $_t(s) = 1$ and $_t(s') = 0$ for all $s' \neq s$. If a patient is in one of the partially observable states, i.e., states NC, $C1$, $C2$, and Re, the belief vector can be represented as $_t = (0, _t(NC), _t(C1), _t(C2), _t(Re), 0, 0, 0)$, where $_t(NC) + _t(C1) + _t(C2) + _t(Re) = 1$.

Immediate Rewards In our CPOMDP, the immediate reward is the QALD. In this paper, we will use the terms *treatment outcome*, *QALDs*, and *health gain* interchangeably. To represent the QALDs, we assign a value $h^a_t(s) \in [0, 1]$, in every period to every action and health state pair (a, s). That is, the immediate reward $r_t(a, s)$, that a patient gains at state s under treatment decision a at time t is $h^a_t(s)$, for $t \in \{0, \ldots, N-1\}$[4]. We denote the terminal reward $r_N(a, s)$, which equals to $h_N(s)$ in our model. Therefore, the belief state immediate health reward under treatment a at time t is $\sum_{s \in S} r_t(a, s) \ _t(s)$, and we denote this value $h^a_t(_t)$. To avoid an overly myopic optimal policy, we calculate the expected health gain for the patient under no treatment from day 360 to day 420 and add it as the terminal reward. We discuss the significance and sensitivity of the terminal reward in § 2.6.4.

[4] In our model the immediate health reward $h^a_t(s)$, is independent of t. That is, $\forall t, h^a_t(s) = h^a(s)$. We include the parameter t here for generality.

Costs The expected cost at time t under treatment a at state s, $c_t^a(s)$, is always negative.[5] When $t \in \{0, \ldots, N-1\}$, $c_t^a(s)$ includes the cost of detoxification if the patient is in the state Dx, the cost of hospital visits if a patient overdoses (see Ÿ 2.9), and the cost of treatment if the patient is not in the states Dx and Abs. Note that there is no terminal expected cost under this formulation, i.e. $c_N^a(s) = 0$ for all states $s \in S$ and actions $a \in A$, since the cost that will incur next year should be separated from the cost that will incur this year. Thus, the belief state expected cost under treatment a at time $t \in \{0, \ldots, N-1\}$ is $\sum_{s \in S} c_t^a(s)\, \beta_t(s)$, and we denote this quantity $y_t^a(\beta_t)$.

Budget Constraint The budget constraint is incorporated via an open-loop optimization formulation. Let $\beta_t > 0$ denote the amount of budget that is allocated for the patient for the rest of the horizon, in month t. In each month t, we solve a new optimize problem using the budget β_t. To reflect the costs of wearable devices, we deduct the cost of the specific wearable from the annual budget before solving the optimization problem.

Optimality Equations The optimal treatment action sequence maximizes the expected reward gained throughout the planning horizon. The optimal solution can be solved using dynamic programming techniques (for example, see Cassandra et al. (1997)). Let $V_t^*(\beta_t)$ denote the optimal expected value of the objective function at time t under belief β_t. Let $P(o_{t+1} \mid a, \beta_t)$ be the probability of making observation o_{t+1} after taking action a at belief state β_t, that is $P(o_{t+1} \mid a, \beta_t) = \sum_{s_{t+1} \in S} w(o_{t+1} \mid s_{t+1}, a) \sum_{s_t \in S} \beta_t(s_t) P(s_{t+1} \mid s_t, a)$. Let β_{t+1} be the updated belief given the old belief β_t, observation o_{t+1}, and action a. Then, the optimal Bellman's equation for an unconstrained POMDP satisfies:

$$V_t^*(\beta_t) = \max_{a \in A} \left\{ \sum_{s_t \in S} r_t(s_t, a)\, \beta_t(s_t) + \sum_{o_{t+1} \in O} P(o_{t+1} \mid a, \beta_t) V_{t+1}^*(\beta_{t+1}) \right\} \tag{2.1}$$

$$= \max_{a \in A} \left\{ \sum_{s_t \in S} \beta_t(s_t)\, r_t(s_t, a) + \sum_{o_{t+1} \in O} \sum_{s_{t+1} \in S} w(o_{t+1} \mid s_{t+1}, a) P(s_{t+1} \mid s_t, a) V_{t+1}^*(\beta_{t+1}) \right\}. \tag{2.2}$$

We express β_{t+1} in terms of known model parameters, that is, the transition probability $P(s_{t+1} \mid s_t, a)$ and observation probability $w(o_{t+1} \mid s_{t+1}, a)$:

$$\begin{aligned} \beta_{t+1}(s_{t+1}) &:= P(S_{t+1} = s \mid o_{t+1}, a, \beta_t) = \frac{P(o_{t+1}, s \mid a, \beta_t)}{P(o_{t+1} \mid a, \beta_t)} = \frac{P(o_{t+1} \mid s, a, \beta_t) P(s \mid a, \beta_t)}{P(o_{t+1} \mid a, \beta_t)} \\ &= \frac{w(o_{t+1} \mid s, a) \sum_{s' \in S} P(S_{t+1} = s, S_t = s' \mid a, \beta_t)}{P(o_{t+1} \mid a, \beta_t)} \\ &= \frac{w(o_{t+1} \mid s, a) \sum_{s' \in S} P(s \mid s', a) \beta_t(s')}{P(o_{t+1} \mid a, \beta_t)}. \end{aligned} \tag{2.3}$$

[5] Similarly, the expected cost $c_t^a(s)$ in our model is also independent of

The second to the last equality is because the observation o_t is independent of the belief b_t. Note that the belief vector in conjunction with the belief update absorb the entire history of the model and thus achieve the Markovian property (Smallwood and Sondik 1973).

Constrained POMDP Model The goal of the constrained POMDP model is to find the optimal treatment plan, $\pi(b_t)$, that yields the maximum expected reward through the planning horizon while satisfying the budget constraint w_t, where Π denotes the set of of all feasible policies. Recall $h^a_t(b_t)$ is the expected health reward of taking treatment a at time t under the clinicians' belief about patient's health state b_t, and let H_0 denote the expected health under an initial belief b_0 and policy π in our CPOMDP problem, i.e.,

$$H_0 = E_0\left[\sum_{t=0}^{N+1} h^a_t(b_t) + h_N(b_t)\right] \tag{2.4}$$

Unless otherwise mentioned, we fix this initial belief b_0 throughout the rest of the paper, and note that H_0 is always non-negative. Similarly, let C_0 be the expected cost under an initial belief b_0 and the policy π in our CPOMDP problem, i.e.

$$C_0 = E_0\left[\sum_{t=0}^{N+1} c^a_t(b_t)\right] \tag{2.5}$$

Note that C_0 is always non-positive. We formulate the following optimization problem:

$$\text{System I:} \quad \pi^*(b_t) = \arg\max_{\pi\in\Pi} H_0 \tag{2.6}$$

$$\text{s.t.} \quad C_0 \le w_t \tag{2.7}$$

Constraint (2.7) guarantees that the expected cost of the treatment will be less than or equal to the budget. Since we model the problem as an open loop problem (we reoptimize the problem using an updated budget at each time step), it is guaranteed that our final solution would satisfy the budget constraint if we could solve the problem using System I.

Let $\{\pi^*(b_t)\}_{b_t\in[0,1]}$ be the set of optimal solutions in System I when we vary b_t from 0 to 1. Let $H_0(\pi^*(b_t))$ denote the expected health under policy $\pi^*(b_t)$. Before discussing the tractability of System I, we first list some properties that this optimal solution set satisfies:

Proposition 1. Properties of the optimal solution set in System I:

1(a) The optimal policy $\pi^*(b_t)$ is not necessarily unique, but the optimal expected health $H_0(\pi^*(b_t))$ is unique for any fixed b_t and b_0.

1(b) The optimal expected health $H_0(\pi^*(b_t))$ is non-decreasing in b_t.

1(c) Let $\ddot{H}_0$ denote the unique elements contained in the optimal solution set $H^*_0(\beta_t)$ for $\beta_t \in [0, \beta_0]$. Then, the set $\ddot{H}_0$ is finite.

Property 1(a) follows directly from uniqueness of Equation (2.6). Property 1(b) holds because the current optimal solution remains feasible after budget increase. Property 1(c) holds because the number of actions we can perform is finite and the initial belief b_0 is fixed.

The objective function, Equation (2.6), can be solved through exact POMDP algorithms using Equations (2.1) (2.2). However, with Constraint (2.7), System I is numerically intractable to solve (Poupart et al. 2015). To address this problem, we provide a novel reformulation of our CPOMDP problem by incorporating the budget constraint into the objective function in the next section, and show that our reformulation can be solved using exact POMDP algorithms in conjunction with a binary search. When randomized policies are allowed, our reformulation solves the original problem exactly, but when deterministic policies are enforced, our reformulation might find a suboptimal solution.

2.5 Analytical Results

In this section we present our reformulation of the CPOMDP problem. The key idea in our budget reformulation is to integrate the budget constraint into the objective function through a tunable parameter, and optimize over this parameter. Mathematically, the immediate reward becomes a convex combination of the expected health and cost in our reformulation. Thus, our immediate reward at time t under treatment a at state s, $r_t(a, s)$, becomes $h^a_t(s) + \lambda * c^a_t(s)$, where $\lambda \in [0, 1]$, $h^a_t(s) \in [0, 1]$ and $c^a_t(s) < 0$, for $t \in \{0, \ldots, N-1\}$, and the terminal reward $r_N(a, s)$, becomes $h_N(s)$. The belief state immediate reward under treatment a, $r_t(a, b_t)$, takes the following form:

$$r_t(a, b_t) = \sum_{s \in S} r_t(a, s) b_t(s) = h^a_t(b_t) + \lambda * c^a_t(b_t),$$

for $t \in \{0, \ldots, N-1\}$, and the belief state terminal reward is $h_N(b_N) = h_N(b_N)$. Let $V(\pi, b_0, \lambda)$ denote the expected reward function under a policy π, initial belief b_0, and parameter λ. By the linearity of expectation, we have the following Lemma (whose proof is included in § 2.10):

Lemma 2. If the immediate reward and terminal reward of a discrete time finite horizon POMDP takes the form $h^a_t(b_t) + \lambda * c^a_t(b_t)$ for every time step, then $V(\pi, b_0, \lambda)$ can be written as $V(\pi, b_0, \lambda) = H(\pi, b_0) + \lambda * C(\pi, b_0)$.

The rest of this section is organized as follows: we first state our CPOMDP reformulation in § 2.5.1 and then show the correctness of our reformulation in § 2.5.2 by comparing the solutions obtained in System I and our reformulation. We will then review incremental pruning, one of the

exact solutions for unconstrained POMDP and show how to use it in conjunction with a binary search to solve our reformulation in Ÿ 2.5.3.

2.5.1 CPOMDP Reformulation

For a xed initial belief $_0$, our budget reformulation comprises the following optimization system:

$$\text{System II:} \quad \pi^< = \arg\max_{\ddot{E}} \{ * C_0 : * C_0 w_t \} \tag{2.8}$$

$$< = \max \{ : * C_0^< w_t; \quad \ddot{E} [0,1] \} \tag{2.9}$$

$$\pi^< = \arg\min_{\ddot{E}^<} H_0 \tag{2.10}$$

$$\pi^< = \arg\max_{\ddot{E}} \{ H_0 + .1 * /C_0 \}; \tag{2.11}$$

and we denote the optimal policy under budget $_t$ in System II by $\pi^<(\ _t)$. System II describes a two-step optimization problem. First, in Equation (2.11), for a xed , we nd the set of optimal policies, $\pi^<$, that nds the maximum expected reward throughout the planning horizon. In Equation (2.10), we pick the $\pi^<$ that has the smallest expected health $H_0^<$. Since the expected cost is negative, Equation (2.10) is also equivalent to nding the $\pi^<$ that yields the largest expected cost $C_0^<$. Second, in Equation (2.9), we nd the largest, $<$, such that the absolute expected cost that we nd in Equation (2.11), $|C_0^<|$, is under the budget, $_t$. After nding $<$, we resolve Equation (2.8) and denote this optimal policy $\pi^<$, inside $\pi^<_<$ with the largest absolute cost such that the cost is under the budget constraint. Thus, we obtain the optimal solution $\pi^<(\ _t)$. The purpose of Equations (2.10) and (2.8) is to handle the situations where and $\pi^<_<$ contain multiple optimal solutions. Since all policies inside $\pi^<_<$ have the same objective function value $V^<(\ _0; \ _t)$, maximizing the expected cost is equivalent to maximizing the expected health in Equation (2.8). In other words, Equation (2.8) is equivalent to $\arg\max_{\ddot{E}\,\pi^<_<} \{ H_0 : * C_0 w_t \}$.

Let $\hat{}(\pi^<(\ _t)\grave{}_{t\ddot{E}[0,\emptyset)}$ denote the set of optimal policies in System II when we vary $_t$ from 0 to Ø, and let $H_0^<(\ _t)$ denote the maximum expected health under policy $\pi^<(\ _t)$. Let $V^<(\ _0; \ _t)$ denote the optimal value of System II under budget $_t$. Let $\pi^<_<$ be the optimal policy with the lowest expected health under the optimal parameter $<$ as de ned in Equation (2.9), i.e., $\pi^<_< = \arg\min_{\ddot{E}\,\pi^<_<} H_0$. Then, similar to Proposition 1, we rst list the properties that the optimal solutions in System II satisfy:

Proposition 2. Properties of the solution set in System II:

2(a) The optimal policy $\pi^<(\ _t)$ is not necessarily unique, but the values $H_0^<(\ _t); C_0^<(\ _t)$ are unique for any xed $_t$.

Figure 2.1: Construction of the e ciency frontier of the original formulation's feasible solution set convex hull

2(b) $H_0^{<}, * C_0^{<}$, and the objective function $H_0^{<} + .1 * /C_0^{<}$ are all non-decreasing in . Furthermore, as $_t$ increases, $^{<}$ is non-decreasing.

2(c) $H_0^{<}. _t/$ is non-decreasing in $_t$.

2(d) If $_2 > _1$ and $V^{<}. _0; _2/ > V^{<}. _0; _1/, H_0^{<}. _2/ > H_0^{<}. _1/$.

2(e) Let $H_0^{<}$ denote the unique elements contained in the set $H_0^{<}. _t/ _{t \in \{0,\emptyset/}$. Then, the set $H_0^{<}$ is nite.

Proposition 2 shows that the optimal solution set in System II satis es the properties of the optimal solutions in System I. Property 2(a) is directly implied by Equation (2.8). We prove Property 2(b) via construction, and the proof is included in Ÿ 2.11.1. Property 2(c) follows from Property 2(b), and the proof can be found in Ÿ 2.11.2. Property 2(d) states that under the same initial belief $_0$, for any budget $_2$ that is su ciently larger than budget $_1$ so as to obtain a strictly higher objective value, the lowest possible expected health among the set of optimal policies, $H_0^{<}. _2/$, is always greater than the highest expected health that we can obtain under budget $_1$, $H_0^{<}. _1/$. Property 2(d) is implied directly by Property 2(c) and also the proof of Property 2(b). After showing the correctness of our reformulation, Property 2(e) follows from Property 1(c) since the solutions that we nd in System II is a subset of that in System I.

Note that our budget reformulation can be applied to any CPOMDP problem with one constraint, e.g., the optimal breast cancer diagnose problem with budget constraint proposed in Ayvaci et al. (2012), or the quickest change detection problem proposed in Isom et al. (2008). In addition, in Ÿ 2.13 we show that our reformulation can be extended to the case where we have multiple constraints. However, the running time of the current solution algorithm might grow exponentially as the number of constraints increases. Novel algorithms to solve the extended reformulation could be proposed, however this is beyond the scope of this paper.

2.5.2 Correctness of Our Reformulation

Proposition 1 suggests that, in System I, $H_0(\cdot\,; t)$ is a non-decreasing step function in t. Figure 2.1 (left) shows a plausible $H_0(\cdot\,; t)$ function, where each solid dot represents a set of optimal solutions that yield the same maximum expected health and the same expected cost which is under the budget t, and each solid line represents that the optimal solution set stays the same over an interval of budget values. Note that in System I it is possible that there are multiple solid dots on each red line segment, i.e., there exists multiple optimal solutions with different expected costs but the same expected health. However, this is not allowed in our reformulation, System II. This is because when there are multiple sets of solution with the same expected health but different expected costs, our objective function, Equation (2.11) will always prefer the solution set with the lowest expected cost.

If we connect the end points of this step function $H_0(\cdot\,; t)$, then we obtain a piecewise linear, non-decreasing function $F(t)$. Fig. 2.1 (middle) illustrates the construction of $F(t)$ defined on $[l; u]$, where each point on the solid line represents a randomize policy that is a convex combination of the two nearest optimal solution sets (i.e., the two nearest solid dots). However, depending on the structure of each individual problem, this function $F(t)$ is not necessarily concave (as in our illustrating example, the middle figure in Fig. 2.1). Let $\hat{F}(t)$ be the point-wise smallest concave function whose hypograph contains the hypograph of $F(t)$. Then, $\hat{F}(t)$ is the efficiency frontier of the convex hull of the original formulation's (System I's) feasible solution set. That is, when randomized policies are allowed, the solutions that lie on $\hat{F}(t)$ are optimal. If $F(t)$ is concave, then $\hat{F}(t) = F(t)$, and if it is not, then when randomized policies are allowed, we will be able to find solutions lying on $F(t)$ that either dominate or are equal to the solutions lie on $\hat{F}(t)$. Fig. 2.1 (right) illustrates the construction of $\hat{F}(t)$, where $F(t)$ is denoted with the solid curves. Indeed, this function $\hat{F}(t)$ is piecewise linear, concave, and strictly increasing.

Next, in Theorem 3 we show that the our formulation, System II, recovers the convex hull of the feasible solution set in System I :

Theorem 3 (Correctness of our reformulation): Let $F(t)$ be piecewise linear, concave, and strictly increasing on $[l; u]$[6] as defined above, then $H_0^* = H_0^<$ when C_0^* lies on $F(t)$. That is, our reformulation, System II finds all solutions that lie on the efficiency frontier of the convex hull of the solutions of System I.

The proof of Theorem 3 can be found in Ÿ 2.12. Theorem 3 implies that System II recovers the convex hull of the feasible solution set in our original formulation. That is, when randomized policies are allowed, our reformulation finds the exact solution to the original problem. However,

[6] We pick $l; u$ such that F is $-\infty$ on $[0; l)$ (i.e., infeasible) and constant on $[u; \infty)$

when deterministic solutions are enforced, our reformulation will miss any optimal solutions that lie strictly within the convex hull of solutions. In our illustrating example, when deterministic policies are enforced, our System II will find the optimal policies that correspond to π_1, π_3, π_4, and π_u, but it will fail to find the optimal policy that corresponds to π_2 in Fig 2.1 (right). Note that Theorem 3 agrees with the findings in Ayvaci et al. (2012) where they show that when randomized policies are allowed, the optimal health outcome is strictly concave with respect to the budget. Although our reformulation provides stronger theoretical guarantees when randomized policies are allowed, to align our results with past literature on clinical decisions, we consider only deterministic optimal policies in our experiments.

2.5.3 Solution Method for Our Reformulation

In System II, we can compute Equations (2.8), (2.10), and (2.11) using the exact POMDP algorithm incremental pruning (Zhang and Liu 1996, Cassandra et al. 1997). To solve for Equation (2.9), we observe that the monotonicity of $H_0^<$ and $C_0^<$ with respect to λ in Property 2(b) implies that the feasible region of $\lambda \in [0,1]$ in Equation (2.9) is continuous, and thus we can solve for u using a binary search in conjunction with any exact solution method for finite-horizon POMDPs. Before we provide a complete algorithm that solves our reformulation, we first review one of the exact finite-horizon POMDP solution methods that we will use, incremental pruning.

Review of Incremental Pruning An Exact POMDP Solution Method The incremental pruning algorithm (Zhang and Liu 1996, Cassandra et al. 1997) can be applied to solve finite-horizon POMDPs exactly using backward induction; it relies heavily on the following decomposition of the optimal Bellman's equation (Equation (2.2)):

$$V_t^{a_t;o_{t+1}}(\pi_t) = \sum_{s_t \in S} \pi_t(s_t) \left[\frac{r(a_t,s_t)}{|O|} + \sum_{s_{t+1} \in S} w(o_{t+1}|s_{t+1},a_t) P(s_{t+1}|s_t,a_t) V_{t+1}^<(\pi_{t+1}) \right]; \tag{2.12}$$

$$V_t^{a_t}(\pi_t) = \sum_{o_{t+1} \in O} V_t^{a_t,o_{t+1}}(\pi_t); \tag{2.13}$$

$$V_t^<(\pi_t) = \max_{a_t \in A} V_t^{a_t}(\pi_t). \tag{2.14}$$

In Equation (2.12), for a fixed belief π_t, we calculate the expected reward for each fixed observation o_{t+1} after we have taken action a_t, and we denote this value $V_t^{a_t;o_{t+1}}(\pi_t)$. In Equation (2.13), we then sum over all possible observations and thus obtain $V_t^{a_t}(\pi_t)$. Lastly, in Equation (2.14), we take the maximum over the set of all possible actions, selecting the action that yields the highest expected reward. One can easily verify that the above decomposition is equivalent to Equation (2.2).

To distinguish the belief state expected values from the expected value at each underlying states $s \in S$ at time t, we introduce α-vectors, α_t, to represent the vector containing the expected

reward at each state s_i. Let $V_N(s_i)$ be the expected reward of state s_i at time N. Then,

$$\alpha_N := [V_N(s_1), \ldots, V_N(s_{|S|})]. \tag{2.15}$$

Note that since we do not make decisions in period N, α_N is independent of actions. Thus, the belief state expected reward becomes $V_N(b) = \sum_{s_N \in S} b(s_N) V_N(s_N) = \alpha_N \cdot b$.

Since the transformations (2.12) (2.14) preserve the piecewise linearity and convexity of the optimal Bellman's equation with respect to the belief b (Sondik 1971, Smallwood and Sondik 1973, Cassandra et al. 1997), there exists some unique finite parsimonious representations of the value functions $V_t^{a_t,o_{t+1}}$, $V_t^{a_t}$, and $V_t^<$ (Sondik 1971, Zhang and Liu 1996). Let $A_t^{a,o}$ denote the set of α-vectors under a fixed action a_t, observation o_{t+1} at time t. Let A_t^a be the set of α-vectors under a fixed action a_t at time t summed over the set of all possible observations $o_{t+1} \in O$ as described in Equation (2.13). Let A_t be the set of α-vectors at time t that includes all actions $a_t \in A$. We will provide the formal definition of those three variables below in Equations (2.20) (2.22). Using induction, for some α-vectorsets $A_t^{a,o}$, A_t^a, and A_t, we can write

$$V_t^{a_t,o_{t+1}}(b) = \max_{\alpha \in A_t^{a,o}} \alpha \cdot b; \tag{2.16}$$

$$V_t^{a_t}(b) = \max_{\alpha \in A_t^a} \alpha \cdot b; \tag{2.17}$$

$$V_t^<(b) = \max_{\alpha \in A_t} \alpha \cdot b. \tag{2.18}$$

Equations (2.16) (2.18) can effectively reduce the number of vectors that we need to keep track of in Equations (2.12) (2.14). So the idea of the algorithm is to find the set of α-vectors that are undominated at every belief state in each backward induction step.

Let $purge(\cdot)$ be the minimal representation of a set by removing the pointwise dominated vectors, and let $\oplus$ be the Minkowski sum of two sets.[7] Let $\alpha_{t+1}(s_{t+1})$ be the expected reward at state $t+1$ with $\alpha_{t+1} \in A_{t+1}$, i.e., $\alpha_{t+1}(s_{t+1}) = V_{t+1}(s_{t+1})$. Then, we define $\alpha(\alpha_{t+1}, a_t, o_{t+1})(s_t)$ to be the scaled expected reward at state s_t after taking action a_t and making the observation o_{t+1}, i.e.,

$$\alpha(\alpha_{t+1}, a_t, o_{t+1})(s_t) = \frac{r(a_t, s_t)}{|O|} + \sum_{s_{t+1} \in S} w(o_{t+1}, s_{t+1}, a_t) P(s_{t+1}, s_t, a_t) \alpha_{t+1}(s_{t+1}). \tag{2.19}$$

Thus, the minimal representation of the sets $A_t^{a,o}$, A_t^a, A_t at time t can be represented by

$$A_t^{a,o} = purge(\{\alpha(\alpha_{t+1}, a, o) | \alpha_{t+1} \in A_{t+1}\}); \tag{2.20}$$

$$A_t^a = purge\left(\bigoplus_{o \in O_{t+1}} A_t^{a,o}\right); \tag{2.21}$$

$$A_t = purge\left(\bigcup_{a \in A_t} A_t^a\right). \tag{2.22}$$

[7] $X \oplus Y = \{x + y : x \in X, y \in Y\}$

Equation (2.20) corresponds to Equation (2.16), where we calculate the set of scaled expected reward vectors for all $|A|$-vectors obtained from the last iteration $\dot{E}\,A_{t+1}$ using Equation (2.19), under a fixed observation o_{t+1} and action a. We then remove all vectors that are pointwise dominated by another vector inside this set as those vectors will never appear in the optimal solution. Similarly, Equations (2.21) and (2.22) correspond to Equations (2.17) and (2.18), where we sum over the set of all observations and combine the $|O|$-vectors under different actions respectively, and obtain the set of undominated vectors. Equation (2.21) can be solved efficiently using incremental pruning (Cassandra et al. 1997), and Equations (2.20) and (2.22) can be implemented efficiently using the algorithm Lark Prune (White 1991). The pseudocode of these two algorithms are provided in Ÿ 2.14).

Algorithm 1 Solving unconstrained POMDP

Solve POMDP: $\langle S, A, T, R, O, \Omega, \gamma, b_0 \rangle$

1: $A_N \leftarrow \{\langle V_N(s_1), \ldots, V_N(s_{|S|}), a_N = \emptyset \rangle\}$. each element in A_N is a tuple
2: **for** t in $[N-1, \ldots, 0]$ **do** . iterate over time backwards
3: **for** a in A_t **do** . iterate over actions
4: **for** o in O_{t+1} **do** . iterate over observations
5: **for** v in A_{t+1} **do** . iterate over alpha-vector sets obtained from the last iteration
6: $u \leftarrow \{v[0], \ldots, v[|S|]\}$. initialize u to be the 1st and 2nd element of v, respectively
7: $U \leftarrow \emptyset$
8: **for** d in $[29]$ **do** . calculate the expected reward and cost in the next 29 days
9: $u = \langle r_t(a, s_1), \ldots, r(a, s_{|S|}) \rangle + P_t^a \cdot u$
10: $= \langle c_t^a(s_1), \ldots, c_t^a(s_{|S|}) \rangle + P_t^a$
11: **end for**
12: U.add(u, a, o, c, γ) . calculate alpha-vector using Equations (2.19,2.23)
13: **end for**
14: $A_t^{a,o} \leftarrow$ Lark Prune(U) . whether $\langle u, a \rangle \in U$ will be pruned depends on u, i.e., reward
15: **end for**
16: $A_t^a \leftarrow$ Incremental Pruning($A_t^{a,o_1}, \ldots, A_t^{a,o_{|O|}}$) . similar to Line 14
17: **end for**
18: $A_t \leftarrow$ Lark Prune($\bigcup_a A_t^a$) . similar to Line 14, and we keep each set A_t^a separate in A_t
19: **end for**
20: $E \leftarrow \arg\max_{\langle v, a \rangle \in A_0} v[0] \cdot b_0$. find all elements in A_0 with maximum expected reward
21: **return** E, E.action . return set E and its associated action set E.action

---

Algorithm 2 Solving System II

Main Solver. $_t$; $_0$/

$_0$ C 0, $_1$ C 1 . keep track of the lower bound and upper bound of the current

C_{old} C *Ø . keeps track of the expected cost from the last iteration

; C C 1 . initialize the change in and the optimal expected cost to some number>

Y_0; Y_0:action C Solve POMDP = 0; $_0$/ . check the extreme points

Y_1; Y_1:action C Solve POMDP = 1; $_0$/

m_0 C arg max$_{y \ddot{E} Y_0}$ y[1] . nd the feasible policy with lowest expected cost

m_1 C arg min$_{y \ddot{E} Y_1}$ y[1] . nd the feasible policy with highest expected cost

if ð$_0$ .m_0[1]/ð> $_t$ then return problem not feasible

else if ð$_0$ .m_0[1]/ð= $_t$ then return m_0:action, $_0$ .m_0[0]/; $_0$ .m_0[1]/

else if ð$_0$ .m_1[1]/ðw $_t$ then return m_1:action, $_0$ .m_1[0]/; $_0$ .m_1[1]/

end if

while > or C > do . the loop stops if both w and C w

C . $_0$ + $_1$/_2 . calculate current value

Y ; Yaction C Solve POMDP; $_0$/ . obtain the set of optimal solutions

m C arg max$_{y \ddot{E} Y}$ y[1] . nd the element with the lowest expected health

if ð$_0$ m[1]ð= $_t$ then break loop . we have found the optimal <

else if ð$_0$ m[1]ð< $_t$ then $_0$ C . if feasible, update the lower bound$_0$ to the current

else $_1$ C . update the upper bound$_1$ to

end if

C . $_1$ * $_0$/, C C ðC$_{old}$ * . m[1]/ $_0$ð . update the change in and C

C_{old} C .m[1]/ $_0$. updateC_{old}

end while

<; Y<, Y<:action C ; Y ; Y.action . we have found < in the above while loop

m< C arg min$_{y \ddot{E} Y<}$^y[1] : y[1] w $_t$` . solve Equation (2.8)

return m<:action, $_0$ m[0], $_0$ m[1]; < . return optimal action, expected reward, cost, and

The Solution to Our Reformulation To compute the expected cost as de ned in System II, similar to the de nition of $_t$, we let $_t$ be a vector denoting the expected cost at each state at time t. Thus, we have

$$c_t(s_t, a_t, o_{t+1}) = \frac{c_t^a}{N} + \sum_{s_{t+1} \in S} w(o_{t+1} \mid s_{t+1}, a_t) P(s_{t+1} \mid s_t, a_t) \, _{t+1}(s_{t+1}). \tag{2.23}$$

Let (u, a, d) and $_c(\cdot, a, o)$ be two -vectors of length , where the i^{th} element of (u, a, d) and $_c(\cdot, a, o)$ are $(u, a, d)(s_i)$ and $_c(\cdot, a, o)(s_i)$, respectively. Algorithm 1 describes the complete algorithm to solve the unconstrained POMDP problem (Equation (2.11) in System II) aligning the transition and reward matrices with our one-month decision period (since the transition and reward matrices are de ned in days). Note that the inputs of the three pruning steps, in Lines (14) (18) of Algorithm 1, are sets of tuples with length 2, and whether a tuple will be pruned in those steps depends on the rst entry of the tuple. That is, the three pruning steps are performed over the rst entry of each element in the input sets.

In order to solve System II, one slight modi cation to the algorithm Lark Pruner is necessary. In the original Lark Pruner algorithm, if there are two policies that give the same optimal objective value, then the tie is broken arbitrarily. However, in our algorithm, we maintain all policies that produce the same optimal objective value and only discard an optimal policy if it admits the same expected health and cost with another policy.

Algorithm 1 includes only the necessary variables to demonstrate the solution methods of System II. For example, to obtain the expected health directly from the algorithm, one could add an additional element h_N towards the end of the tuple A_N in Line 1 and update the rest of the algorithm accordingly. Similarly, we could add variables to keep track of the expected health coming from the terminal reward. Furthermore, this structure allows us to plug in di erent transition matrices P_t^a in Lines (9) (12) when comparing di erent models. We will discuss this in more detail in our numerical section, Ÿ 2.6.

Algorithm 2 solves our reformulation System II, where a binary search over the space is implemented to nd the optimal $<$ value given the budget constraint and an initial belief state. To obtain randomized policies, one could discretize the space of and nd the convex hull of the optimal solution set. The complexity of our algorithm is exponential in the numbers of states and actions, and polynomial in the parameter . In the case where all states are fully observable, i.e., we have a MDP, Algorithm 2 can be modi ed to incorporate any exact solution methods for nite-horizon MDPs, and the complexity of our algorithm is polynomial in the numbers of states and actions, and .

Note that since the costs are always negative, the expected cost at a state

$$_t(s) = \sum_{o \in O_{t+1}} \; _c(\cdot, a, d)$$

is non-increasing as t decreases in Algorithm 1 for all action a and state s. Thus, to reduce the run-time of our algorithm, in addition to pruning (Lines 14-18), we could remove an element in the set A_t if the expected cost evaluated at t exceeds the budget. While our reformulation can be solved optimally within a reasonable time window when problem size is comparatively small (as in our problem), we provide a heuristic algorithm (with a worst-case error bound) in Ÿ 2.15 to further reduce the run-time.

2.6 Numerical results

In this section, we numerically investigate the bene t of incorporating wearable devices in OUD treatments using the reformulation and Algorithm 2 presented in Ÿ 2.5. We are collaborating with a startup[8] to test the feasibility of di erent wearables (Fitbit, Garmin, and Empatica E4) in capturing patients' craving states: thirty patients from Jade Wellness Center in Pennsylvania are participating in this pilot study. One initial observation is that craving frequencies di er greatly across patients[9], con rming the need to study the bene ts of a personalized treatment strategy. However, since the data that we have collected thus far is insu cient for accurate parameter estimation, the majority of the parameters of our CPOMDP model were estimated from past literature. (We can con rm that these are in the range of the observed pilot data.) For those that were not available from the literature, we perform extensive sensitivity analysis. Our model is written in Python and takes an average of 6,025 seconds to run all 7 cases (Case 1, Case 2a, Case 2b, Case 2c, Case 2d, Case 3, and the benchmark case) to completion on a 3.2 GHz Core i7-8700 machine with 64 GB Ram. All cases took on average 14 iterations to terminate.

In Ÿ 2.6.1 we describe the various types of wearable devices. In Ÿ 2.6.2 we describe our transition matrices and parameter estimation. In Ÿ 2.6.3 we describe our numerical results. We discuss sensitivity analysis and case extensions in Ÿ 2.6.4. We denote the average transition probabilities of the entire patient population as the average treatment dynamics (ATD).

2.6.1 Comparison of Cases

Table 2.2 summarizes all cases that we study. Case 1 contains only monthly urine tests (cost $360 per year), while Cases 2 and 3 additionally study three types of wearables (and two hypotheticals). The rst type costs $120 (Fitbit); the second type costs $258 (Garmin); the third type costs $1200

[8] We provided them with a rst set of wearables. This startup subsequently received funding by the NSF to complete their SBIR Phase I study.

[9] Also observed in various past literature (Chalana et al. 2016). A more detailed discussion on factors that could a ect patient treatment response can be found in Ÿ 2.2 and Ÿ 2.6.2.

Case Num.	Technology	Device Cost	Device Accuracy	Device Life Span
1	UT	-	-	-
2a	wearables with UT	$120	0.6	1
2b	wearables with UT	$258	0.7	1.5
2c	wearables with UT	$500	0.8	2
2d	wearables with UT	$800	0.9	2.5
3	wearables	$1200	1	3
4	benchmark	0	1	-

Table 2.2: Costs associated with di erent cases. In Cases 1, 2, and 3, we assume the urine tests all cost $360 per year. The device accuracy is measured in its sensitivity and speci city in detecting patients' craving episodes, and the device life spans are measured in years. UT stands for urine test

(Empatica E4). Recall that in Ÿ 2.4, we de ned the sensitivity (q_s) and speci city (p_s) for every partially observable state $s \in S$, for each type of wearable. Because our experimental results are robust under small perturbations in q_s and p_s, we set $q_s = p_s$ for all partially observable states s, and $q_s = q_{s'}$ for every partially observable states pair $s; s'$. Henceforth, we refer to the value of q_s associated with each wearable type as the device accuracy. According to the number of features that each type of wearable collects, we set the device accuracy to be 0.6, 0.7, and 1 for types 1, 2, and 3 wearables, respectively, where a device accuracy of 0.5 is equivalent to guessing randomly. To ll out intermediate values, we consider two additional hypothetical types of wearables: accuracy of 0.8 and cost $500, and accuracy of 0.9 and cost $800. Because costlier wearables have longer life spans, we spread the price over their life spans. We obtain our benchmark case, Case 4, by setting the cost of the wearable devices and urine test in Case 3 to zero.

Wearables can also help to determine patient response to di erent treatments (personalized treatment dynamics, PTD) in Cases 2, 3, and 4 (there is no information in Case 1 that will allow us to do that). In our experiments, we compare the scenarios where PTD is equal to or divergent from ATD. In addition, we consider three levels of patient treatment adherence levels, where high, medium, and low TAs represent that the patient follows the treatment above 90%, between 70% and 90%, and below 70% of the time, respectively.

2.6.2 Transition Probability and Parameter Estimation

Transition Probability Matrices Recall that $P(s_{t+1} \mid s_t; a)$ is the transition probability from state $s_t \in S$ to state $s_{t+1} \in S$ under action $a \in A$, where t indicates the number of days since the last drug-use. Let P_t^a be the daily transition probability matrix in which the rows correspond to

and columns to s_{t+1}. Since patients' treatment adherence levels can affect the evolution of their conditions and hence the transition probabilities, having different treatment adherence levels could potentially affect the additional health benefits brought by wearable devices. Thus, $R_t^a(s)$ is also a function of a patient's treatment adherence level.

Although past literature also suggests that other variables including the patient's age, jail history, comorbid medical conditions, drug injection history, number of past overdoses, and number of relapses could also affect the transition probabilities, we assume that theses variables are automatically captured in the resulting transition matrices. In addition, since the horizon in our model is short (at most twelve months), we assume that the numbers of past overdoses and relapses remain the same (regardless of the possibility that the patient might reach states OD or Re in the rest of the horizon), to ensure the tractability of our model. Since the budget is modeled as an open-loop optimization problem (i.e. at the beginning of each month, we solve a new optimization problem with an updated budget and the patient's information), if the patient relapsed within a month, then the transition matrices will first be updated in the following month to incorporate this information.

Omitting t and a, let $x_s; nc_s; c_s^2; e_s$, and od_s be the probability of transitioning to Dx, NC, C2, Re, and OD, from health state s, respectively; let d_s be the probability of dying at state s and w be the probability of withdrawing from any state with any action at any time.[10] We represent the transition probability matrices as follows:

$$P_t^a = \begin{array}{c} \text{from\_to} \\ \text{Dx} \\ \text{NC} \\ \text{C1} \\ \text{C2} \\ \text{Re} \\ \text{OD} \\ \text{Dt} \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Dx} & \text{NC} & \text{C1} & \text{C2} & \text{Re} & \text{OD} & \text{Dt} & \text{Abs} \end{array} \\ \begin{pmatrix} x_{Dx} & nc_{Dx} & Z_1 & c_{Dx}^2 & 0 & od_{Dx} & d_{Dx} & w \\ 0 & nc_{NC} & Z_2 & 0 & 0 & 0 & d_{NC} & w \\ 0 & nc_{C1} & Z_3 & c_{C1}^2 & e_{C1} & od_{C1} & d_{C1} & w \\ 0 & 0 & Z_4 & c_{C2}^2 & e_{C2} & od_{C2} & d_{C2} & w \\ x_{Re} & 0 & Z_5 & c_{Re}^2 & e_{Re} & od_{Re} & d_{Re} & w \\ 1 * d_{OD} & 0 & 0 & 0 & 0 & 0 & d_{OD} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

where $Z_1 = 1 * x_{Dx} * nc_{Dx} * c_{Dx}^2 * od_{Dx} * d_{Dx} * w; Z_2 = 1 * nc_{NC} * d_{NC} * w; Z_3 = 1 * nc_{C1} * c_{C1}^2 * e_{C1} * od_{C1} * d_{C1} * w; Z_4 = 1 * c_{C2}^2 * e_{C2} * od_{C2} * d_{C2} * w; Z_5 = 1 * x_{Re} * c_{Re}^2 * e_{Re} * od_{Re} * d_{Re} * w$.

Throughout our paper, we assume that if a patient reaches any of the states OD, Dt, or Abs, clinicians will be notified immediately; that is, these states are always observable. Fig. 2.2 depicts the transition diagram between our health states.

[10] The assumption that w is independent of states and action is observed by Termorshuizen et al. (2005).

Figure 2.2: Transition diagram. States C1 and C2 share the same types of transitions, however the transition probabilities need not to be the same.

According to Chalana et al. (2016), if patients relapse right after detoxi cation, they will overdose with high probability due to the loss of tolerance. After experimenting with various values, we nd that our results are robust with respect to the probability transitioning from Dx to Rehab, e_{Dx}, and thus we set it to be 0 (i.e. a patient will overdose with probability 1 if relapsing the day after detoxi cation). Moreover, in our transition matrices, the probability of transitioning to state OD decreases as t increases. We assume that patients either die or receive treatments followed by detoxi cation if they overdose, and that patients must experience some craving the day before relapse if they are not in state Dx.

Parameter Estimation We initialize two sets of transition matrices $\hat{T}_0^{a'}$ and $\hat{T}_1^{a'}$. The former is estimated from past literature (Ÿ 2.16) and re ects how a patient reacts to di erent treatments on average, i.e., the ATD. The latter matrix is randomly perturbed around the ATD to represent the true treatment dynamics of a given patient (patient ground truth). Therefore, to nd the optimal policy, we set the transition matrices to be $\hat{T}_0^{a'}$ in Case 1 (since we do not have access to wearables and thus could not develop PTD), and set them to be $\hat{T}_1^{a'}$ in Cases 3 and 4. In Case 2, since the wearables provide imperfect information, we assume that the accuracy of the estimated transition probability matrices increases as the accuracy of the wearable increases. Thus, we randomly perturb the transition matrices $\hat{T}_1^{a'}$ in Case 2 with the perturbation magnitude decreasing as the wearable accuracy increases.[11] We refer to the resulting expected health and cost as the estimated (or observed) expected health and cost respectively.

To evaluate and compare the expected health in each case, we use $\hat{T}_1^{a'}$ as our transition probability matrices and the identity matrix as our observation matrix to calculate the expected health, and we refer to these as the true expected health and cost respectively. Past literature shows that the treatment IN is superior to treatment M and B; therefore, we assume in our experiment

[11] Thus the information our algorithm uses is a more accurate estimate of the true transition matrices.

that P_t^{IN} is stochastically equal to or better than P_t^M and P_t^B. The e ects of TA are estimated from Nosyk et al. (2009). In the rest of this section, unless otherwise stated, expected health refers to the true expected health

In our experiments, we consider a patient with the following characteristics: the patient has relapsed 5 times, overdosed once, and has just completed a 21-day detoxi cation program. (We experimented with di erent patient characteristics, and the results are similar.) We assume that the initial belief state is high craving intensity (C2), and that an average patient reacts to treatment B better than to treatment M. To avoid overly myopic solutions, we calculate the expected health gain for the patient under no treatment from day 360 to day 420, weight it by γ, and add it as the terminal reward.

2.6.3 Results

In our baseline scenario (Ÿ 2.6.3), we examine the bene t of adopting wearables in OUD treatments in the case where the patient reacts to treatments exactly as the average treatment dynamics predict, implying that, by de nition, there is no bene t of implementing PTD. We consider three budget levels: low ($9K), medium ($15K), and high ($21K). When the budget is low, we could not a ord MAT in every single period; when the budget is medium, we could a ord MAT is all periods and can a ord to treatment IN occasionally; when the budget is high, we could a ord treatment IN in the majority of periods. Finally, we explore two possibilities within PTD: when the patient reacts to treatment M better than to B (Ÿ 2.6.3) and when the patient reacts to treatment B better than to M (Ÿ 2.6.3).

Since all cases contain monthly urine tests, we assume the state relapse is observable (at the end of the month). Thus, in this section, we have only three partially observable states: NC, C1, and C2, and all cases are solved using the exact algorithm. Unless stated otherwise, our results are robust with respect to terminal rewards. Because our budget formulation nds only the points that lie on the e ciency frontier of the health-budget curve, and we consider only deterministic optimal policies in our numerical experiments below (as randomized policies may be problematic in the eld), we sometimes observe that our optimal solution does not spend all the budget. The cost gap, the gap between the observed expected cost and our budget, indicates how close our solution is to the optimal solution when the problem is fully observable. We de ne the value of wearables to be the di erence between the expected health when we incorporate wearables and that of Case 1.

Figure 2.3: The Baseline Scenario. Left: 9K budget with medium TA. Middle: 15K budget with medium TA. Right: 21K budget with medium TA. At termination, the optimal λs are approximately [0.053, 0.039, 0.077] (left), [0.294, 0.269, 0.274] (middle), and [0.301, 0.303, 0.310] (right), where the number of digits that $<$ contains equals to the number of iterations that took System II to terminate; the expected QALDs received from treatments (excluding the terminal reward) are [285.44, 280.98, 289.58] (left), [313.0, 316.03, 317.9] (middle), and [326.18, 329.49, 311.36] (right).

Baseline scenario

In this scenario, the patient reacts to di erent treatments in keeping with the average treatment dynamics. Fig. 2.3 (left top) shows the true expected QALD gains under the optimal policy for di erent cases. In Fig. 2.3 (left bottom), the estimated expected cost is represented by the triangle; the true expected costs of di erent elements are represented by the colored bars. Since the improvement in health between the benchmark model and Case 1 is limited, in Fig. 2.3 we omit Case 2 for ease of illustration. As we will see later, the values of Case 2 wearables are often dominated by that of Case 3.

When the budget is \$9K (Fig. 2.3 left), the value of the wearable is negative. This is because the bene t of observing health states with higher certainty is undermined by the fact that we are left with less money to treat the patient. As the budget increases to \$15K and \$21K (Fig 2.3 middle and right), the value of the wearables is positive but remains negligible: the health state relapse is observable and the rewards for di erent craving states are similar, so the bene ts of wearables are limited. Further experiments suggest that this observation holds for patients with di erent TA levels and di erent transition matrices. In addition, we observe in Fig. 2.3 that for each case, as the budget increases, the optimal value indeed increases as indicated by Proposition 2(b). In this section, we conclude that if all patients react to treatments in accordance with the average treatment dynamics that is, individual variability is low, then we should not adopt wearables in OUD treatments.

Figure 2.4: Scenario 1 where the patient reacts to treatment M better than B. Top row: 9K budget. Middle row: 15K budget. Bottom row: 21K budget. Left column: medium TA. Middle column: high TA. Right column: low TA. At termination, from top left to bottom right, the optimals are approximately [0.052, 0.039, 0.030, 0.043, 0.037, 0.036, 0.040], [0.406, 0.115, 0.205], [0.099, 0.055, 0.056], [0.294, 0.187, 0.135, 0.121, 0.106, 0.102, 0.103], [0.534, 0.335, 0.339], [0.126, 0.058, 0.058] [0.300, 0.222, 0.181, 0.165, 0.145, 0.133, 0.141], [0.551, 0.419, 0.432], and [0.146, 0.060, 0.060].

Scenario 1

In this scenario, the patient reacts to treatment M better than to B, and PTD is implemented in Cases 2, 3, and 4. We rst assume that the patient is allocated with a $9K budget and has medium TA. We observe in Fig. 2.4 (top left) that the expected health increases as the accuracy of the wearable device increases. In addition, the cost gap decreases signi cantly as the device accuracy increases. In Fig. 2.4 (top middle) we observe that the value of wearable devices decreases signi cantly. This is because the high adherence causes the outcome of less expensive treatment (for example, treatment M) to become comparable to that of the most expensive treatment, treatment IN. Similarly, Fig. 2.4 (top right) shows that the value of wearables decreases when comparing to Fig. 2.4 (top middle), this time because patients with poor treatment adherence react poorly to both treatments M and B. When comparing gures in Fig. 2.4 top row, we observe that when the budget is low, the expected costs of overdose and detoxi cation increases as the patient treatment adherence level decreases. This is because patients with lower adherence levels are more likely to relapse when the treatment is inconsistent, i.e., when we cannot a ord to provide treatment in every single period.

Now, if we increase the budget from $9K to $15K, in Fig. 2.4 (middle left), we nd that the value of more expensive wearables increases slightly when compared with Fig. 2.4 (top left) Although the values of wearables remain positive when budget is increased to $21K in Fig. 2.4 (bottom left) they are reduced as the more expensive treatment is utilized more frequently. Again, when TA is high or low, we observe in Fig 2.4 (middle middle and right, bottom middle and left) that the value of the wearable is less than that when TA is medium. Moreover, the value of the wearables slightly decreases as the budget increases because we can a ord treatment IN more frequently. Overall, the value of wearables increases when patients' PTD vary from the ATD.

Scenario 2

In this scenario, the patient reacts to di erent treatments in the same order as the ATD (i.e., reacts to treatment B better than M), but with a di erent magnitude. Because many patterns we observe in this scenario are similar to those in Scenario 1, we highlight the di erences. We observe that in Fig. 2.5 (middle), the value of the wearable becomes negative when compared with Fig. 2.4 (top middle). In Fig. 2.5 (left) because the cost of the urine test and wearable device is high in Case 3 when compared to our budget, although the increase in treatment outcome between Cases 4 and 1 is relatively large, the health improvement is relative minor in Case 3. In Fig. 2.5 (right) when the patient has low treatment adherence, we observe that the value of wearables increases slightly when compared with that of Scenario 1 in Fig. 2.4 (top right). In addition, this value is similar to the one observed in Fig. 2.5 (top left) When we increase the budget in Scenario 2, we observe in Fig. 2.7 (Ÿ 2.17) that wearables can still improve patient treatment outcomes. Finally, in Ÿ 2.6.3 and

Figure 2.5: Scenario 2 where the patient reacts to treatment B better than M but with a di erent magnitude than in the average treatment dynamics. Budget level equals to 9K. TA equals to medium, high, and low from left to right. At termination, from left to right, the optimal 's are approximately [0.052, 0.034, 0.025, 0.025, 0.025, 0.027, 0.029], [0.406, 0.026, 0.106], and [0.099, 0.052, 0.053]; the expected QALDs received from treatments (excluding the terminal reward) at termination are [200.73, 202.68, 200.19, 200.97, 172.13, 206.40, 224.61], [297.29, 293.30, 301.72], and [125.33, 135.47, 152.74].

Ÿ 2.6.3, we observe that when the accuracy of the wearables is low, we often observe a large gap between the observed and actual expected cost and thus either undertreat the patients or exceed the budget constraint.

2.6.4 Sensitivity Analysis and Case Extensions

Sensitivity to the Cost of Urine Tests Recall that the cost of urine tests is $360 per year, which is higher than the cost of a wearable with device accuracy less than 0.8. On the other hand, if the wearable has high device accuracy, then monthly urine tests become less valuable. Thus, in our experiments, we also explored the possibility of having less frequent urine tests for Case 2 (while the state relapse remains fully observed). Similar to Ÿ 2.6.3 and Ÿ 2.6.3, we observe wearables with low to medium device accuracies often underestimate the costs (e.g., in Fig. 2.8 in Ÿ 2.17). When we compare Fig. 2.8 with Fig.s 2.4, 2.5, and 2.7, we nd that having less frequent urine tests often increases the health outcome even when the device accuracy is low. In contrast, having less more frequent urine tests is bene cial when device accuracy is low.[12]

[12]Recall that the actions in our model can only be changed once per month, and we assume that the urine test result is obtained right before considering the change of treatment in the next period.

frequent urine tests always increases the health outcome when device accuracy is high.

Sensitivity to Other Model Parameters To check the robustness of our results, we experimented with di erent mechanisms to perturb the transition probability matrices and observe similar results with a few exceptions: 1) the magnitude of the value of wearables when the patient has low treatment adherence and reacts to treatment M better than B is sensitive to di erent matrix perturbation mechanisms (e.g., see Fig. 2.9 (left) in Ÿ 2.17); 2) when the device accuracy is medium, we sometimes observe that the value of wearables could also be negative (e.g., see Fig. 2.9 (right) in Ÿ 2.17) because the increase in the cost of the wearables exceeds the bene t of the additional information gathered by the wearables. Both observations highlight the importance of knowing the exact device cost-accuracy tradeo s, and could be potential research topics for future eld studies.

We found that the rest of our insights are robust under small perturbations in the transition matrices, as long as the stochastic order of patient reaction to di erent treatments remains the same. This robustness also holds for immediate rewards, costs, and the initial belief state. However, our model can be sensitive to the terminal rewards. In particular, if the terminal reward is 0 and the budget is low, our optimal solution becomes myopic, in the sense that the patient might end up in bad health states in the end of the horizon with high probability. On the other hand, one should also be cautious when setting a high terminal reward, as the optimal policy might be changed: intuitively, when the terminal reward is too large, we will treat patients with the most expensive treatment towards the end of the horizon to try to gather the terminal reward.

Extensions We extended our cases to incorporate EMA survey devices where daily survey responses are collected from patients. In particular, we consider the possibility that the patient may provide truthful, random, or falsi ed information. We describe the model for this extension in Ÿ 2.18. In this case the transition probability matrices could be directly estimated from the survey results. Thus, the accuracy of the transition probability matrices would decrease when the patient provides noisy or falsi ed information more frequently. While the value of EMA survey devices dominates that of wearables when completion frequency is high and the patient is completely honest, we see that in our eld study that patients or their surrogates do not answer surveys very regularly, thus reducing the feasibility of EMA devices. At the extreme, the value of EMA device can be negative when the patient provides only noisy information.

2.7 Summary of Findings and Concluding Remarks

In this work, we have built a POMDP model with budget constraints to evaluate the potential value of incorporating di erent wearables in OUD treatments. We provide a novel budget reformulation and show that it can be solved e ciently.

Our experiments with our model, incorporating parameters calibrated to literature, show that value of wearables increases as the patient treatment dynamics deviate from the ATD. In addition, as the budget increases, the value of wearable devices eventually decreases when compared to our baseline model. This is because when the budget is relatively low, choosing the cost-e ective treatment could potentially increase patient treatment outcome by a large margin. When the budget is su ciently high, treatment IN guarantees the same treatment outcome for all patient treatment adherence groups and is superior to all other treatments. The value of wearables thus becomes negligible.

We also discover that wearables are more bene cial to patients with medium to low treatment adherence levels; patients with high treatment adherence react to cheaper treatments better than those with lower treatment adherence, and it becomes more cost-e ective to treat the patient even without the wearables. Similarly, we observe that when the budget is xed, as the patient treatment adherence decreases, the patient is more prone to relapse; thus, incorporating wearables can e ectively reduce the risk of relapse. Not surprisingly, as the budget increases, the marginal return of money spent on treatments or wearables decreases the fastest for patients with high treatment adherence.

These observations imply that when the budget that we have is large enough to treat every patient with treatment IN, there is no bene t of adopting wearables. On the other hand, if the budget is smaller, to maximize the treatment outcome for the entire patient group, we should prioritize patients with higher treatment adherence; in this case, the bene t of incorporating wearables in OUD treatments is also limited. However, in the situation where we have enough budget to treat every patient but cannot a ord treatment IN for everyone, to maximize the treatment outcome for the entire patient group, we should allocate more money to patients with lower treatment adherence and incorporate wearables for those patients. Although maximizing treatment outcome for the entire patient population is a reasonable objective, the ethics behind which patient we should allocate resources to is debatable. This question, however, is beyond the scope of this paper.

Future work includes collecting data from planned eld studies in collaboration with rehab centers and start-ups that are developing wearable devices for tackling OUD to ne tune our models and analysis. Throughout the paper, we have assumed that the use of wearables in OUD treatments does not a ect patient TA levels as well as treatment outcomes. It would also

be interesting to see if there are behavioral changes attributable to wearables, even if we only passively collect data, i.e., not sending warnings/reminders, as not all types of wearables have this capability. We hope our work here adds to the growing literature and motivates further work to tackle this important issue.

2.8 Case 1 Parameter Calculation

Let $30C1$ be the event of experiencing $C1$ on day 30. Then

$$c1_{ut+} = P.ut + ã 30C1/ = \frac{P.ut + ã 30C1/}{P.30C1/} = \frac{P.ut + ã 30C1/}{\sum_{i=1}^{8} ..f_t^{a}/^{30}/_{i;3}\_8};$$

where $..f_t^{a}/^{30}/_{i;3}$ is the i^{th} row and 3^{rd} entry of $.f_t^{a}/^{30}$. Because all rows of $f_t^{a}/^{30}/_{i;3}$ sum up to 8, which is the number of our states, we renormalize the value of $P(C1$ on day 30) such that it is between 0 and 1. Furthermore,

$$P.ut + ã 30C1/ = P.30C1 \& 27Re/P.27Re/ + P.30C1 \& 28Re/P.28Re/ + P.30C1 \& 29Re/P.29Re/$$

where $27Re$, $28Re$, and $29Re$ are defined the same way as $30C1$, that is, experiencing relapse on day 27, 28, and 29, respectively. For example, $P.30C1 \& 27Re/ = ..f_t^{a}/^{3}/_{5;3}$ and $P.27Re/ = \sum_{i=1}^{8} ..f_t^{a}/^{27}/_{i;5}\_8$

2.9 Values of Immediate Rewards and Costs

States	Health
Dx	0.6*
NC	1£
C1	0.9
C2	0.75
Re	0.683
OD	0.1*
Dt	0

Abs	withdraw the program from state Dx:	0.5• remaining horizon*
	from state NC:	1 • remaining horizon
	from state C1:	0.9• remaining horizon
	from state C2:	0.75• remaining horizon
	from state Re:	0.678• remaining horizon
	from state Dx:	0.5• remaining horizon

Table 2.3: Health reward associated with each state. *: assumptions that we made based on conversations with clinicians. In sensitivity analysis, we found that our results are robust under small perturbations of those parameters. The rest of the values are generated according to Connock et al. (2007)£: the health reward of state NC is the baseline utility.

Item	Cost type	Costs
Urine tests	model cost	$30/month (Schackman et al. 2012)
Wearable devices	model cost	see Table 2.2
EMA survey device messaging cost	model cost	$60/year (Model in Appendix 2.18)
Physician and nursing time	treatment cost	$50/month (Schackman et al. 2012)
Methadone maintenance treatment	treatment cost	$20/month (Barnett 2009)
Buprenorphine maintenance treatment	treatment cost	$200/month (Barnett 2009)
Implant Naltrexone (Vivitrol)	treatment cost	$1200/month
Counseling	treatment cost	$600/month
OD	state cost	$2500/episode*
Detoxi cation	state cost	$1500/episode

Table 2.4: The costs considered in the model. The values without references in the table were determined through conversations with clinicians from the perspective of the healthcare system. *: the expected cost of overdose includes the expected cost of hospitalization and emergency room visit.

Action/Treatment	Costs
No treatment	0
Methadone maintenance treatment with counseling	$670/month
Buprenorphine maintenance treatment with counseling	$850/month

Implant Naltrexone with counseling	$1800/month
Counseling	$600/month

Table 2.5: Action associated costs. This table summarizes the cost of taking each action using the values provided in Table 2.4.

2.10 Proof of Lemma 2

Lemma 2. If the immediate reward and terminal reward of a discrete time finite horizon POMDP takes the form $h_t^a(\cdot\,,\,\cdot\,) + \lambda * c_t^a(\cdot\,,\,\cdot\,)$ for every time step, then $V(\cdot\,,\,\cdot\,;\,\lambda)$ can be written as $V(\cdot\,,\,\cdot\,;\,\lambda) = H(\cdot\,,\,\cdot\,) + \lambda * C(\cdot\,,\,\cdot\,)$.

Proof. Proof of Lemma 2 Let N be the length of the horizon. We will proceed by induction, and show that at each step t, the value function of an action a_t and belief β_t, $V_t^{a_t}(\cdot\,,\,\beta_t;\,\lambda)$, can be transformed into

$$V_t^{a_t}(\cdot\,,\,\beta_t;\,\lambda) = h_t^{a_t}(\cdot\,,\,\beta_t) + \lambda * c_t^{a_t}(\cdot\,,\,\beta_t). \tag{2.24}$$

First, recall that the value function of an action a with belief β_t of a general POMDP is

$$V_t^a(\cdot\,,\,\beta_t) = \sum_{s_t \in S} r_t(s_t, a)\,\beta_t(s_t) + \sum_{o_{t+1} \in O} P(o_{t+1} \mid a, \beta_t)\, V_{t+1}^{<a,o>}(\cdot\,,\,\tau(\beta_t, a, o_{t+1})).$$

Note first that our immediate reward at time t under treatment a at state s has the form $r_t(a, s) = h_t^a(s) + \lambda * c_t^a(s)$, where $\lambda \in [0, 1]$, $h_t^a(s) \in [0, 1]$ and $c_t^a(s) < 0$, for $t \in \{0, \ldots, N-1\}$, and our terminal reward at state s at time N is $r_N(a, s) = h_N(s) + \lambda * c_N(s)$.[13]

Base case: Equation (2.24) holds when $t = N-1$. First, at $t = N$ for a fixed λ, the value function is

$$V_N(\cdot\,,\,\beta_N;\,\lambda) = \sum_{s_N \in S} h_N(s)\,\beta_N(s) + \lambda * c_N(s)\,\beta_N(s) = h_N(\cdot\,,\,\beta_N) + \lambda * c_N(\cdot\,,\,\beta_N).$$

Then, at $t = N-1$ with action a and belief β_{N-1}, the value function of our POMDP satisfies

$$\begin{aligned} V_{N-1}^a(\cdot\,,\,\beta_{N-1};\,\lambda) &= \sum_{s \in S_{N-1}} r_{N-1}(s, a)\,\beta_{N-1}(s) + \sum_{o \in O_N} P(o \mid a, \beta_{N-1})\, V_N(\cdot\,,\,\tau(\beta_{N-1}, a, o);\,\lambda) \\ &= h_{N-1}^a(\cdot\,,\,\beta_{N-1}) + \lambda * c_{N-1}^a(\cdot\,,\,\beta_{N-1}) + \sum_{o \in O_N} P(o \mid a, \beta_{N-1}) \cdot \left(h_N(\cdot\,,\,\beta_N) + \lambda * c_N(\cdot\,,\,\beta_N) \right) \\ &= H_{N-1}^a(\cdot\,,\,\beta_{N-1}) + \lambda * C_{N-1}^a(\cdot\,,\,\beta_{N-1}); \end{aligned}$$

[13] Note that in our problem $c_N(s) = 0$, but we want to prove our theorem under the most general setup.

where $H^a_{N*1}(\cdot\ _{N*1})$ and $C^a_{N*1}(\cdot\ _{N*1})$ are the expected health and cost at time $N*1$ under action a and belief $_{N*1}$, respectively:

$$H^a_{N*1}(\cdot\ _{N*1}) = h^a_{N*1}(\cdot\ _{N*1}) + \sum_{o \in O_N} P(o|a;\ _{N*1})H(_N(\cdot\ _N); \text{ and}$$

$$C^a_{N*1}(\cdot\ _{N*1}) = c^a_{N*1}(\cdot\ _{N*1}) + \sum_{o \in O_N} P(o|a;\ _{N*1})c(_N(\cdot\ _N):$$

Induction step: to show Equation (2.24) holds for any step t, we assume it holds for step $t+1$, and show that it also holds for step t.

$$\begin{aligned} V^a_t(\cdot\ _t;\ ) &= \sum_{s \in S} r_t(a; s)\ _t(s) + \sum_{o \in O_{t+1}} P(o|a;\ _t)V^{a_{t+1}}_{t+1}(\cdot\ _{t+1};\ ) \\ &= h^a_t(\cdot\ _t) + (1*\ )c^a_t(\cdot\ _t) + \sum_{o \in O_{t+1}} P(o|a;\ _t)(H^{a_{t+1}}_{t+1}(\cdot\ _{t+1}) + (1*\ )C^{a_{t+1}}_{t+1}(\cdot\ _{t+1})) \\ &= H^a_t(\cdot\ _t) + (1*\ )C^a_t(\cdot\ _t); \end{aligned}$$

where $H^a_t(\cdot\ _t)$ and $C^a_t(\cdot\ _t)$ are the expected health and cost at time t under action a and belief $_t$, respectively:

$$H^a_t(\cdot\ _t) = h^a_t(\cdot\ _t) + \sum_{o \in O_t} P(o|a;\ _t)H^{a_{t+1}}_{t+1}(\cdot\ _{t+1}); \text{ and}$$

$$C^a_t(\cdot\ _t) = c^a_t(\cdot\ _t) + \sum_{o \in O_t} P(o|a;\ _t)c^{a_{t+1}}_{t+1}(\cdot\ _{t+1}):$$

. □

2.11 The Remaining Proofs of Proposition 2

Let's first recall Proposition 2:

Proposition 2. Properties of the solution set in System II:

2(a) The optimal policy $^<(\cdot\ _t)$ is not necessarily unique, but the values $H^<_0(\cdot\ _t); C^<_0(\cdot\ _t)$ are unique for any fixed $_t$.

2(b) $H^<_0, *C^<_0$, and the objective function $H^<_0 + (1*\ )C^<_0$ are all non-decreasing in . Furthermore, as $_t$ increases, $^<$ is non-decreasing.

2(c) $H^<_0(\cdot\ _t)$ is non-decreasing in $_t$.

2(d) If $_2 > \ _1$ and $V^<(\cdot\ _0;\ _2) > V^<(\cdot\ _0;\ _1)$, $H^<_0(\cdot\ _2) > H^<_0(\cdot\ _1)$.

2(e) Let $H^<_0$ denote the unique elements contained in the set $H^<_0(\cdot\ _t)\ _{t \in \{0,1,2,\ldots\}}$. Then, the set $H^<_0$ is finite.

2.11.1 Proof of Property 2(b)

Recall by Lemma 2, at each time t, for a fixed belief π_t, the optimal Bellman's equation for our POMDP (Equation (2.11)) at time t and belief state π_t can be written as a convex combination of the expected health and expected cost:

$$V_t^<(\pi_t, \lambda) = \max_{a \in A_t} \{(1-\lambda) * H_t^a(\pi_t) + \lambda * C_t^a(\pi_t) + \sum_{o \in O_{t+1}} P(o|a, \pi_t) V_{t+1}^<(\pi_{t+1}, \lambda)\}$$

$$= \max_{a \in A_t} \{(1-\lambda) * H_t^a(\pi_t) + \lambda * C_t^a(\pi_t)\},$$

where $H_t^a(\pi_t)$ and $C_t^a(\pi_t)$ are the expected health and cost at time t under action a and belief π_t (as defined in § 2.10). Let $V_t^a(\pi_t, \lambda) = (1-\lambda) * H_t^a(\pi_t) + \lambda * C_t^a(\pi_t)$. Note that $H_t^a(\pi_t)$ and $C_t^a(\pi_t)$ are fixed under the optimal value function from the last iteration $V_{t+1}^<(\pi_{t+1}, \lambda)$, current belief π_t, and action a, where $V_{t+1}^<(\pi_{t+1}, \lambda)$ is treated as a constant in the the optimal Bellman's equation. To ease notation, for a fixed π_{N*1}, let $a^<(\pi_{N*1}, \lambda)$ be the action that yields the highest expected reward at time $N*1$ for some fixed parameter λ, that has the lowest expected health:

$$a^<(\pi_{N*1}, \lambda) = \arg\min_{a \in \arg\max_{a \in A_{N*1}} V_{N*1}^a(\pi_{N*1}, \lambda)} H_{N*1}^a(\pi_{N*1}).$$

It suffices to show that at time $t = N*1$, for every fixed π_{N*1}, the first part of Property 2(b) holds, that is,

Claim 1. as λ increases, $H_{N*1}^{a^<(\pi_{N*1}, \lambda)}(\pi_{N*1}, \lambda) * C_{N*1}^{a^<(\pi_{N*1}, \lambda)}(\pi_{N*1}, \lambda)$, and $V_{N*1}^{a^<(\pi_{N*1}, \lambda)}(\pi_{N*1}, \lambda)$ are non-decreasing in λ.

This is because 1) the optimal Bellman's equation at any time step can be written in the same format as $t = N*1$, and 2) the exact backward induction algorithms in solving POMDP relies on fixing the belief state at every searching point. (See § 2.5.3 for detailed discussion on one of the exact POMDP solution method)

Before proving Claim 1, we first introduce some notation and a key concept in the proof the effective action set. Since we fix the belief π_{N*1} for the rest of the proof, we abbreviate $H_{N*1}^{a^i}(\pi_{N*1})$ with H_{a^i}, abbreviate $C_{N*1}^{a^i}(\pi_{N*1})$ with C_{a^i}, and abbreviate $V_{N*1}^a(\pi_{N*1}, \lambda)$ with $V(a, \lambda)$. Under a fixed belief state π_{N*1}, given a set of actions at time $N*1$, A_{N*1}, we define the effective action set A_{N*1} to be the largest subset of A_{N*1} such that all actions lie on the efficiency frontier of the health-cost curve. That is, for every pair of actions in a^1 and a^2 in A_{N*1}, if $H_{a^1} < H_{a^2}$, then $* C_{a^1} < * C_{a^2}$. In other words, we remove all actions that yield equal or worse treatment outcome but have higher cost when compared with another action in set A_{N*1}. For a given λ, let $G_{N*1}(\lambda)$ denote the set of optimal solutions that satisfy $G_{N*1}(\lambda) = \arg\max_{a \in A_{N*1}} V(a, \lambda)$. Then, the actions that yield the largest $V(a, \lambda)$ value are contained in A_{N*1}.

Lemma 3. $\arg\max_{a \in A_{N+1}} V(a, \theta) = G_{N+1}(\theta)$.

Proof. Proof of Lemma 3 For any fixed $\theta \in [0,1]$, let a^1 and a^2 be two actions in set A_{N+1}. If $C_{a^1} > C_{a^2}$, and $H_{a^1} \geq H_{a^2}$, then $a^1 \notin A_{N+1}$, and we must have $V(a^1, \theta) < V(a^2, \theta)$. Thus, action a^1 will never appear in the optimal solution. □

Recall that $a^*(\theta) = \arg\min_{a \in G_{N+1}(\theta)} H_a$. Given a set of effective actions A_{N+1}, without loss of generality, we sort the actions in A_{N+1} according to their expected health from the smallest to the largest, and rename those actions from $1, \ldots, m$, where m is the cardinality of the set A_{N+1}. That is, $H_1 < H_2 < \ldots < H_m$. Thus, proving the first part of Claim 1 is equivalent to showing that $a^*(\theta)$ is non-decreasing as θ increases; this is because as θ increases, $H_{a^*(\theta)}$ increases, and by the definition of the effective action set, $C_{a^*(\theta)}$ also increases. To complete the proof of the first part of Property 2(b), it suffices to show the following propositions:

Proposition 3. The optimal objective function value $V(a^*(\theta), \theta)$, is non-decreasing with respect to θ.

Proposition 4. For a fixed $N+1$, given a sorted effective action set A_{N+1} as indexed above, then the function $a^*(\theta)$ is non-decreasing with respect to θ.

Proof. Proof of Proposition 3 Because $H_i > C_i$, for a fixed action a, the value of $V(a, \theta)$ is strictly increasing in θ. Thus, for every $\theta_i > \theta_j$, $\max_a V(a, \theta_i) = V(a^*(\theta_i), \theta_i) > V(a^*(\theta_j), \theta_i) \geq V(a^*(\theta_j), \theta_j) = \max_a V(a, \theta_j)$. □

Before proving Proposition 4, we will first introduce another key concept that we will use throughout the rest of proof. Given two pairs of points (H_i, C_i) and (H_j, C_j) in the effective action set, with $i < j$, we define $\theta^<_{i,j} = \frac{C_i - C_j}{H_j - H_i + C_i - C_j}$, where $\theta^<_{i,j} \in (0,1)$ is obtained by solving the equation $(H_i + \theta \cdot C_i) = (H_j + \theta \cdot C_j)$ for θ. Note that $\theta^<_{i,j}$ is well-defined when $i < j$ because by construction of the effective action set, $H_j > H_i$ and thus $C_j > C_i$. By the construction of $\theta^<_{i,j}$, we have:

Lemma 4. when $\theta \in [0, \theta^<_{i,j})$, $V(i, \theta) > V(j, \theta)$. Similarly, when $\theta \in (\theta^<_{i,j}, 1]$, $V(i, \theta) < V(j, \theta)$.

To complete the proof of Proposition 4, we need the following corollary implied by Lemma 4:

Corollary 1. If we have a sequence of θs satisfying $\theta^<_{x_1,x_2} \geq \theta^<_{x_2,x_3} \geq \ldots \geq \theta^<_{x_{m-1},x_m}$ with $x_k < x_{k+1}$, $\forall k \in \{1, 2, \ldots, m-1\}$ where $x_k \in A_{N+1}$, $\forall k$, then $a^*(\theta)$ is non-decreasing with respect to θ.

Proof. Proof of Proposition 4 We will proceed by construction. Given a sorted effective action set A_{N+1}, we can find the solution to $V(a, \theta)$ by adding one action from A_{N+1} at a time starting from the smallest indices. We show constructively that adding an action with higher health does not change the monotonicity of $a^*(\theta)$.

Figure 2.6: Proof of Proposition 4 step 2 case II.

Step 0: create an empty list L to keep track of the list of $\preceq$ that are needed to select the optimal action given an input x.

Step 1: add $\preceq_{1,2}$ to L: by Lemma 4, we have monotonicity.

Step 2: add action 3 into the first two actions $\{1, 2\}$, and calculate $\preceq_{2,3}$.

Case I $\preceq_{1,2} < \preceq_{2,3}$: $a^*(x) = 1$ if $x \in [0, \preceq_{1,2}]$, $a^*(x) = 2$ if $x \in (\preceq_{1,2}, \preceq_{2,3}]$, and $a^*(x) = 3$ if $x \in (\preceq_{2,3}, 1]$. We add $\preceq_{2,3}$ to L. Note that at $\preceq_{i,j}$ with $i < j$ we choose action i by the definition of $a^*(x)$, that is, when there are actions with equivalent objective function under the same x, we always pick the one with smaller expected health.

Case II $\preceq_{2,3} < \preceq_{1,2}$ (see Figure 2.6): when $x < \preceq_{2,3}$, we have $V(2, x) > V(3, x)$, and $V(1, x) > V(2, x)$. This implies that $V(1, x) > V(2, x) > V(3, x)$. Thus, when $x \le \preceq_{2,3}$, $a^*(x) = 1$. Similarly, when $x > \preceq_{1,2}$, we have $V(2, x) > V(1, x)$, and $V(3, x) > V(2, x)$. This implies that $V(3, x) > V(2, x) > V(1, x)$. Therefore, when $x \ge \preceq_{1,2}$, $a^*(x) = 3$. Lastly, when $\preceq_{2,3} < x < \preceq_{1,2}$, we have $V(3, x) > V(2, x)$ and $V(1, x) > V(2, x)$. This implies that $a^*(x) = 1$ or 3. Thus, we calculate $\preceq_{1,3}$. In this case, we add $\preceq_{1,3}$ to L and remove $\preceq_{1,2}$ from L (see Figure 2.6).

Case III $\preceq_{1,2} = \preceq_{2,3}$: $a^*(x) = 1$ if $x \in [0, \preceq_{1,3}]$ and $a^*(x) = 2$ if $x \in (\preceq_{1,3}, 1]$. We do not change L. The reasoning is the same as that in Case I.

In all cases, we have monotonicity.

Step i, $i \ge 3$: add action $i + 1$ to the previous actions and calculate $\preceq_{i,i+1}$. Let $\preceq_{j,i} \in L$ for some $j \in \{1, 2, \ldots, i\}$. Note that this index j is in unique in L by construction.

Case I $\preceq_{i,i+1} > \preceq_{j,i}$: when $x \le \preceq_{i,i+1}$, $a^*(x)$ is monotone by previous steps; when $x > \preceq_{i,i+1}$, $a^*(x) = i$. We add $\preceq_{i,i+1}$ to L.

Case II $\preceq_{i,i+1} < \preceq_{j,i}$: then using the same logic as in step 2, we calculate $\preceq_{j,i+1}$, add it to L and remove $\preceq_{j,i}$ from L. We will now check whether $\preceq_{j,i+1}$ violates the condition we defined in Corollary 1. We perform this procedure iteratively for all elements in L until no violation can be found. This procedure terminates in finite time because the number of elements in L is finite, and by construction, when it terminates, we obtain an L that satisfies Corollary 1.

Case III $\alpha^{<}_{i,i+1} = \alpha^{<}_{j,i}$: similar to Case III in step 2, we do not modify the set L, and / remains monotone according to the previous step.

. □

Remark 1. In the above we have shown that for every xed belief state, the rst part of Property 2(b) holds. Just to reiterate, this su ces to show Property 2(b) because the exact backward induction algorithms used in solving a POMDP relies on xing the belief state at every searching point. (See Ÿ 2.5.3 for detailed discussion on one of the exact POMDP solution methods.) Note that since an MDP is a special case of POMDP with the belief state being standard unit vectors, this proof also applies to MDPs.

So far, we have showed that as increases, the value of the objective function $H_0(\alpha^{<}) + .1 * \alpha/C_0(\alpha^{<})$, the minimum optimal expected health $H_0(\alpha^{<})$, and the minimum absolute expected cost $C_0(\alpha^{<})$ are non-decreasing. The proof of Proposition 4 further implies that $C_0(\alpha^{<})$ increases if and only if increases. To complete the proof of Property 2(b), it remains to show that as increases, the optimal $\alpha^{<}$ is non-decreasing. To observe this, we let and α_2 be two budget levels with $\alpha_1 > \alpha_2$. Because the set of all feasible policies under budget is also feasible under budget, Equations (2.9) (2.11) imply that the set of feasible under budget α_2 is also feasible under budget. Let α^1 and α^2 be the optimal $\alpha^{<}$ under budgets α_1 and α_2 respectively. Then, we must have $\alpha^1 \le \alpha^2$ by Equation (2.9).

2.11.2 Proof of Property 2(c)

Proof. To show that $H_0(\alpha^{<}_t)$ is non-decreasing in t, we proceed by contradiction. Let $t_1 > t_2$, and let the optimal policies of System 2/ under $= t_1$ and $= t_2$ be α^1 and α^2, respectively. Let α^1_1 and α^2_1 be the optimal in Equation (2.9) under α^1 and α^2, respectively. Property 2(b) implies that since $t_1 > t_2$, $\alpha^1_1 \le \alpha^2_1$. For the sake of contradiction, we assume $H_0(\alpha^1) < H_0(\alpha^2)$. Since α^2_1 is a feasible solution when $\alpha_t = t_1$, Equation (2.11) implies that

$$t_1 H_0(\alpha^1) + .1 * \alpha^1_1/C_0(\alpha^1) > t_1 H_0(\alpha^2) + .1 * \alpha^2_1/C_0(\alpha^2): \tag{2.25}$$

Since we assume $H_0(\alpha^1) < H_0(\alpha^2)$, we must have $C_0(\alpha^1) \le C_0(\alpha^2)$, that is $* C_0(\alpha^1) w * C_0(\alpha^2)$. In other words, policy α^1 is cheaper than policy α^2 but yields worse outcomes. This implies that α^1_1 is also a feasible solution when $\alpha_t = t_2$. Again, by Equation (2.11), when budget equals to t_2, we have that

$$t_2 H_0(\alpha^1) + .1 * \alpha^1_2/C_0(\alpha^1) < t_2 H_0(\alpha^2) + .1 * \alpha^2_2/C_0(\alpha^2): \tag{2.26}$$

Comparing Equations (2.25) and (2.26), we must have $\alpha^2_2 \ge \alpha^1_1$. Thus we have reached a contradiction.

. □

2.12 Proof of Theorem 3

Theorem 3 (Correctness of our reformulation). Let $F(t)$ be piecewise linear, concave, and strictly increasing on $[l, u]$[14] as defined above, then $H'_0 = H^<_0$ when $*C_0$ lies on $F(t)$. That is, our reformulation, System II, finds all solutions that lie on the efficiency frontier of the convex hull of the solutions of System I.

Since the initial belief $_0$ is fixed throughout this section, for a policy , we will abbreviate H_0 with H and abbreviate C_0 with C. Because F is concave and strictly increasing on $[l, u]$, F is bijective. Thus, $F^{-1}(H_i)$ is well-defined and $F^{-1}(H_i) = *C_i$. Before we prove Theorem 3, we first show the following proposition:

Proposition 5. Let $^<_{i,j} = \frac{C_i * C_j}{H_j * H_i + C_i * C_j}$. If F is piecewise linear, concave, and strictly increasing, and $*C_j$ is a non-differentiable point of F, then $^<_{i,j} < {}^<_{j,k} \forall\ _i, _j, _k$, s.t. $*C_i < *C_j < *C_k$. In particular, there exists $^<_j$ s.t. $^<_j H_j + .1 * {}^<_j / C_j > {}^<_j H + .1 * {}^<_j / C$ for all policy q_j.

We observe that $^<_{i,j}$ is simply an equivalence point of $_i$ and $_j$ in the objective function $H + .1 * /C$ in System 2/; that is, $^<_{i,j} H_i + .1 * {}^<_{i,j} / C_i = {}^<_{i,j} H_j + .1 * {}^<_{i,j} / C_j$. Furthermore, if $> {}^<_{i,j}$, then $^<_{i,j} H_i + .1 * {}^<_{i,j} / C_i < {}^<_{i,j} H_j + .1 * {}^<_{i,j} / C_j$, and if $< {}^<_{i,j}$, then $^<_{i,j} H_i + .1 * {}^<_{i,j} / C_i > {}^<_{i,j} H_j + .1 * {}^<_{i,j} / C_j$.

Proof. Proof of Proposition 5 Because $F^{-1}(H_i) = *C_i$, we rewrite $^<_{i,j} = \frac{C_i * C_j}{F(*C_j) * F(*C_i) + C_i * C_j}$. We first notice that $^<_{i,j}$ is well-defined for all $_i, _j$ if $*C_i > *C_j$, and $^<_{i,j} \in [0, 1]$. Now,

$$
\begin{aligned}
{}^<_{i,j} < {}^<_{j,k} \quad &\Leftrightarrow \quad \frac{C_i * C_j}{F(*C_j) * F(*C_i) + C_i * C_j} < \frac{C_j * C_k}{F(*C_k) * F(*C_j) + C_j * C_k} \\
&\Leftrightarrow \quad \frac{F(*C_j) * F(*C_i) + C_i * C_j}{C_i * C_j} > \frac{F(*C_k) * F(*C_j) + C_j * C_k}{C_j * C_k} \\
&\Leftrightarrow \quad \frac{F(*C_j) * F(*C_i)}{C_i * C_j} > \frac{F(*C_k) * F(*C_j)}{C_j * C_k}.
\end{aligned}
$$

F is piecewise linear, concave, and strictly increasing, and $*C_j$ is a non-differentiable point; therefore, we have

$$
\frac{F(*C_j) * F(*C_i)}{C_i * C_j} \in \left[F'_-(*C_j), F'_+(*C_j) \right][15] \times \frac{F(*C_k) * F(*C_j)}{C_j * C_k}.
$$

Because this holds for all $_i, _k$ s.t. $*C_i < *C_j < *C_k$, without loss of generality, let $*C_i$ be the non-differentiable point right before $*C_j$ and $*C_k$ be the next non-differentiable point after $*C_j$;

[14] We pick l, u such that F is $*\infty$ on $[0, l)$ (i.e., infeasible) and constant on (u, ∞).

[15] $F'_-(x) := \lim_{t \to 0^-} \frac{F(x+t) - F(x)}{t}$; $F'_+(x) := \lim_{t \to 0^+} \frac{F(x+t) - F(x)}{t}$

then for any $\pi \in \Pi(s_i, a_j; s_j, a_k)$ satisfies $H_j^{\pi} + .1 * C_j^{\pi} > H_j^{\pi'} + .1 * C_j^{\pi'}$ for all policy $\pi' \neq \pi$. □

Proof. Proof of Theorem 3 Let π^* be an optimal policy in System II under budget C^{π^*} such that C^{π^*} corresponds to a non-differentiable point on $F(t)$ in System I on the X-axis. To prove Theorem 3, it suffices to show that π^* is the optimal policy in System I when the budget equals C^{π^*}. In other words, we want to show that Equations (2.9)-(2.11) find all non-differentiable points that lie on $F(t)$. Once we have established this, Equation (2.8) will ensure that we find all solutions that lie on the $F(t)$ and are differentiable. To complete the proof, we will proceed by contradiction.

First, the solution pair (H^{π^*}, C^{π^*}) is unique. For the sake of contradiction, assume that when the budget equals C^{π^*}, there exists another policy f s.t. $C^f = C^{\pi^*}$. If $H^f > H^{\pi^*}$, then we have $H^f + .1 * C^f > H^{\pi^*} + .1 * C^{\pi^*}$. This contradicts that π^* is the optimal policy in System II. If $H^f < H^{\pi^*}$, then $H^f + .1 * C^f < H^{\pi^*} + .1 * C^{\pi^*}$ for all λ. Thus, policy f is always dominated by policy π^* and cannot be an optimal solution.

Second, we want to show $H^{\pi} = H^{\pi^*}$. To show $H^{\pi} \in (H^{\pi^*}, H^{\pi^*})$, we observe that all optimal policies in System II: $\{\pi : C^{\pi} \in w_t\}$, are feasible in System I. To show that $H^{\pi} \notin (H^{\pi^*}, H^{\pi^*})$, we again proceed by contradiction. Note that all feasible policies in System I are still feasible in System II; however, since feasibility does not imply optimality in System II, we need to argue more carefully. Assume there exists an optimal solution of System I, π, that satisfies $H^{\pi} > H^{\pi^*}$.

Case I $C^{\pi} \in w C^{\pi^*}$: since π is also feasible in System II, we have $H^{\pi} + .1 * C^{\pi} > H^{\pi^*} + .1 * C^{\pi^*}$. This contradicts that π^* is the optimal policy.

Case II $(t \in C^{\pi} > C^{\pi^*})$: Because π^* is a optimal solution to $\arg\max H^{\pi} + .1 * C^{\pi}$, we must have $H^{\pi^*} + .1 * C^{\pi^*} \geq H^{\pi} + .1 * C^{\pi}$. Because $H^{\pi} > H^{\pi^*}$ and $C^{\pi} > C^{\pi^*}$, Proposition 5 implies that there must exists[16] $f = \lambda_1 \pi + \lambda_2 \pi^*$,[17] with $\lambda_1 > 0$ s.t. $H^{\pi^*} + .1 * C^{\pi^*} < H^f + .1 * C^f$. This contradicts that $\pi^* = \max\{\pi : C^{\pi} \in w_t\}$. □

2.13 Extension to Multiple Constraints

Kim et al. (2011) extends the definition of CPOMDP proposed in Isom et al. (2008) to multiple constraints, where a CPOMDP is defined as a POMDP with two additional components: 1) $c_k(s, a) < 0$ is the cost of type k incurred for executing action a in state s, and 2) $\lambda_k > 0$ is the upper

[16] the existence is implied by the second statement in Proposition 5.

[17] $f \in \Pi$ is implied by the first statement in Proposition 5.

bound on the absolute cumulative cost of type k. We can formally state these constraints as

$$E\left[\sum_{t=1}^{N} c_k(s_t, a_t) \mid w_t\right] \leq \hat{k}.$$

Recall that when we have only one constraint, we reformulated our CPOMDP as follows:

$$\begin{aligned}
\text{System II:} \quad \pi^* &= \arg\max_{\pi \in \Pi} \{R_0 : C_0 \leq \hat{C}_0\} \\
&= \max_{\pi \in \Pi} \{R_0 - \lambda C_0 : \lambda \in [0, \lambda_{\max}]\} \\
&= \arg\min_{\pi \in \Pi} H_0 \\
&= \arg\max_{\pi \in \Pi} \{H_0 + \lambda(\hat{C}_0 - C_0)\}.
\end{aligned}$$

We first illustrate how to incorporate two different types of cost constraints using this reformulation. The extension to the case where $K > 2$ is straightforward. To ease notation, we will omit the initial belief b_0. Let H denote the expected reward under policy π. Let C_1 and C_2 denote the expected cost of type 1 and type 2 under policy π, where C_1 and C_2 are always negative. We denote the upper bound on the absolute cumulative type 1 cost to be $\hat{c}_1$, and similarly $\hat{c}_2$ is the upper bound on the absolute cumulative type 2 cost. Furthermore, let $\lambda_1, \lambda_2 \in [0, \lambda_{\max}]$ be two tunable parameters corresponding to the types 1 and 2 constraints, respectively. Then the idea is to solve two separate problems using System II, where the last steps in our budget reformulations are

$$\begin{aligned}
\pi^* &= \arg\max_{\pi \in \Pi} \{\lambda_1(\hat{c}_1 - C_1)\}; \\
\pi^* &= \arg\max_{\pi \in \Pi} \{\lambda_2(\hat{c}_2 - C_2)\};
\end{aligned}$$

respectively. To solve our reformulation, we perform the following two steps:

Step 1 We solve the problem by pretending only constraint 1 exists (using System II) and find the corresponding optimal parameter λ_1^* and optimal solution π_1^*. Similarly, solve the problem using only constraint 2 and obtain λ_2^* and π_2^*. If one of those solutions also satisfies the other constraint, then we terminate and return that policy.

Step 2 If neither of the optimal solutions π_1^* and π_2^* are feasible, then we need to increase[18] the values of λ_1 and λ_2 iteratively by a small amount until we find a policy that satisfies both constraints with the smallest possible values for λ_1 and λ_2.

[18] Note that in this new formulation, we have flipped the role of λ and $1 - \lambda$ in System II: we need to increase the value of λ here instead of decrease.

The correctness of this extension follows directly from the correctness of our algorithm. However, instead of performing a binary search over the value of when we only have one constraint, we now need to search over the grid of possible values of pairs in Step 2 above. As the number of constraints grows, the running time of our algorithm will grow exponentially with respective to the number of constraints in the problem, and the problem might become computationally intractable. A smarter search mechanism could be proposed in this scenario (e.g., using ideals from the EM algorithm, gradient descent, or bi-objective optimization with -constraints (Mavrotas 2009)), however this is beyond the scope of this problem.

2.14 Algorithms

Algorithms 3-5 are adopted from Cassandra et al. (1997). Let e_s be the standard basis vector that has 1 on the s^{th} entry and 0 everywhere else.

Algorithm 3 Lark's algorithm for purging a set of vectors

Lark Prune(A)

```
1:  F ← A                                                    . dirty set
2:  Q ← ∅                                                    . clean set
3:  for s in S do
4:      λ ← arg max_{v∈F} e_s · v
5:      Q ← Q ∪ {λ}
6:      F ← F \ {λ}
7:  end for
8:  while F is not empty do
9:      for v in F do
10:         x ← Donimate(v,Q)
11:     end for
12:     if x = donimtaed then
13:         F ← F \ {v}
14:     else
15:         λ ← arg max_{v∈F} x · v
16:         Q ← Q ∪ {λ}
17:         F ← F \ {λ}
18:     end if
19: end while return Q
```

Algorithm 4 An LP approach to finding an information state in a vector's witness region

Dominate(, A)

```
L ← LP.variable: x_1, ..., x_|S|; objective: max ε
for ν in A − {μ} do
    add constraint: L, x · (μ − ν) ≥ ε
    add constraint: L, x · 1 = 1
end for
if Infeasible(L) then return dominated
else
    (x, ε) ← SolveLP(L)
    if ε > 0 then return x
    else return dominated
    end if
end if
```

Algorithm 5 Incremental pruning

IncrementalPruning($V_{a,o_1}, \ldots, V_{a,o_k}$)

```
W ← LarkPrune(V_{a,o_1} ⊕ V_{a,o_2})
for i in [3 : k] do
    W ← LarkPrune(W ⊕ V_{a,o_i})
end forreturn W
```

Algorithm 6 Group $A_t^{a,o}$ into subsets according to d_v

regroup($A_t^{a,o}$, d_v)

```
F ← A_t^{a,o}                          . dirty set; each element in F is a tuple of the form (γ, c)
Q ← array of *1 with length(F)         . store the group index that each element in F belongs to
q_index ← 0                            . initialize group index for Q
for i in range(len(F)) do
    if Q[i] ≠ *1 then Continue         . we have already assigned a group index to this element
    else Q[i] ← q_index
    end if
    for j in range(i, len(F)) do
        if Q[j] ≠ *1 then Continue
        else if ‖Q[i].v − Q[j].v‖_1 < d_v then
            Q[j] ← q_index
        end if
    end for
    q_index += 1                       . increment the group index
end for
return Q, q_index                      . return the group index of each element in F, and the total number of groups
```

2.15 Heuristic Algorithm for Solving Unconstrained POMDP

Algorithm 2, incorporating the removal of α-vectors that exceed the budget constraint, can be solved within several hours when the size of the problem is comparatively small. While a few heuristic methods have been proposed to speed up the run-time of finite-horizon POMDPs (Walraven and Spaan 2019), in this section we provide a heuristic algorithm to speed up the running time of our CPOMDP, Algorithm 2, when the complexity of the problem that we are trying to solve increases. This heuristic algorithm relies on reducing the size of the α-vector set, $A_t^{a,o}$, described in Equation (2.20) in the pruning step. Note that: (i) the model described in Ÿ 2.4 of this paper can be solved exactly, and (ii) this heuristic algorithm could also be applied to unconstrained discrete-time finite-horizon POMDPs.

First, we observe that in our problem, when the size of the set $A_t^{a,o}$ is large, the time needed to solve our CPOMDP increases, often because the absolute element-wise difference between some α-vectors is very small. Since the sets A_t^a and A_t are obtained by performing set operations on $A_t^{a,o}$, reducing the size of $A_t^{a,o}$ can also effectively reduce the size of A_t^a and A_t. Thus, the idea of our heuristic algorithm is to remove the α-vector tuples that are "too close" to each other

in the objective vector. We use the L1-norm as our distance metric, but it can be replaced with any p-norms for $p \geq 2$ in practice. In addition, since for a fixed vector x its p-norms, $\|x\|_p$, is monotonically decreasing with respect to p (Rassouli and Jebril 2010), our theoretical guarantee below holds for all p-norms where $p \geq 2$.

Let m_v be a positive tunable parameter that control the threshold distance between two vectors. Algorithm 7 describes the heuristic that we use to reduce the size of $A_t^{a,o}$, where $|D_p|$ is the number of all possible observations that we can have in a partially observable state. The function regroup$(A_t^{a,o}, d_v)$ clusters the elements in $A_t^{a,o}$ so that any two vectors within the same cluster have an L1 distance that is smaller than or equal to d_v (see § 2.14, Algorithm 6). Algorithm 7 can be used inside the incremental pruning (Algorithm 5 in § 2.14) to reduce the running time (from over 24 hours to 20 minutes on randomly generated instances).

Algorithm 7 Reduce the size of the $A_t^{a,o}$ set

reduce$(A_t^{a,o}, m_v, |D_p|)$

```
F ← A_t^{a,o}                          . dirty set; each element in F is a tuple of the form (γ, c)
d_v ← m_v · 2|D_p|                     . initialize the cuto distance for the objective function
Q, total_group ← regroup(A_t^{a,o}, d_v)   . Q store the group index that each element in F belongs to
U ← ∅                                  . clean set;
for g in range(total_group) do
    M ← elements in F that are assigned to group g in Q
    add a random element in M to U
end for
return U
```

First, let V^* be the optimal value of the objective function of an unconstrained POMDP problem (Equation (2.11) of our reformulation) solved using the exact solution, and $\hat{V}$ be the solution obtained by applying Algorithm 7 to Line (14) in Algorithm 1. Let l be the number of horizons that we apply Algorithm 7 to Algorithm 1 in Algorithm 2. Note that $l \leq N$, where N is the length of horizon. First, we show that the objective value that we obtain in Equation (2.11) by applying our heuristic algorithm (Algorithm 7) is at most $m_v \cdot l$ away from the optimal solution:

Theorem 4. $V^* - \hat{V} \leq m_v \cdot l$.

The proof of Theorem 4 can be found in § 2.15.1; this result holds for all unconstrained discrete-time, finite-horizon POMDPs.

2.15.1 Proof of Theorem 4

Theorem 4. $V^* - \hat{V} \leq m_v \cdot l$.

Proof. Proof of Theorem 4 Recall that $V^<$ and $\tilde{V}$ are the optimal values of an unconstrained POMDP problem solved using the exact solution and by applying reduce$(A_t^{a,o}, m_v, 1)$, respectively. Let V' be the optimal value of the POMDP problem solved by applying our heuristic algorithm, Algorithm 7, to Line (18) of Algorithm 1, i.e., reduce$(A_t, m_v, 1)$. We first show that $V^{<*} \geq V' \geq m_v \cdot 1$ by induction, and then show that $V^{<*} - \tilde{V} \geq V^{<*} - V'$.

Base case $t = 1$: note that since the set A_N contains only a singleton according to Equation (2.15), without loss of generality, assume that we reduce the size of A_t at $t = N * 1$, that is, in the Second To the Last Step (STLS) of the backward induction. Let s_{N*1} be the s^{th} entry of the α-vector that we removed from A_{N*1} but is used to compute $V^<$ in the exact solution, and $\alpha_{N*1}(s)$ be the value that is used to compute V'. By construction, we have $\sum_{s \in S} b(s)^<_{N*1}(s) * \alpha_{N*1}(s) \geq m_v$. This implies that $b(s)^<_{N*1}(s) * \alpha_{N*1}(s) \geq m_v$. Recall that by Equation (2.19), at time t equals to $N * 2$, any element inside $A^{a_{N*2},o_{N*1}}_{N*2}$ has the following form:

$$g^{a_{N*2},o_{N*1}}_{\alpha_{N*1}}(s) = \frac{r(a_{N*2},s)}{|O|} + \sum_{s' \in S} \alpha_{N*1}(s') P(s'|s,a_{N*2}),$$

where $\frac{r(a_{N*2},s)}{|O|}$ is a model specific constant, and α_{N*1} is some vector inside A_{N*1}. Let $\alpha^{<a}_{N*2}, \ldots, \alpha^{a}_{N*2} \in A^a_{N*2}$ be the α-vectors that are used to compute $V^<$ and V', respectively, in the set A^a_{N*2} at time $N * 2$. Let A_{N*1} denote the set of α-vectors that we obtained after applying Algorithm 7 to the set A_{N*1}. Thus, we have $\alpha^{<a}_{N*2} \in A_{N*1}$ and $\alpha^{a}_{N*2} \in A_{N*1}$. By Algorithm 1, $\alpha^{<a}_{N*2}$ and α^{a}_{N*2} are obtained by some summing over the set of all possible observations over $\alpha_{N*1}, a_{N*2}, o_{N*1})$, for some α-vectors $\alpha_{N*1} \in A_{N*1}$. Abusing the notation a bit, let $\{x^<_i\}_i$ and $\{y_i\}_i$ denote two sequences of α-vectors inside the set A_{N*1} and A_{N*1}, respectively, where the index corresponds to some observation o_i. (Note that $x^<_i$ could equal $x^<_j$ for $i \neq j$.) In particular, those two sequences of α-vectors were used to generated $\alpha^{<a}_{N*2}$, and α^{a}_{N*2} respectively, i.e., $\alpha^{<a}_{N*2}(s) = \sum_{i=1}^{|O|} g^{x^<_i}(s, a, o_i)(s)$ and $\alpha^{a}_{N*2}(s) = \sum_{i=1}^{|O|} g^{y_i}(s, a, o_i)(s)$. Because $0 \leq w(o_i|s', a) P(s'|s, a) \leq 1$, $\sum_{s' \in S} P(s'|s, a) = 1$, and $\sum_{i=1}^{|O|} w(o_i|s', a) = 1$, we can bound the absolute difference between $\alpha^{<a}_{N*2}(s)$ and $\alpha^{a}_{N*2}(s)$ as follows:

$$\begin{aligned}
\left| \alpha^{<a}_{N*2}(s) * \alpha^{a}_{N*2}(s) \right| &= \left| \sum_{s' \in S} \sum_{i=1}^{|O|} g^{x^<_i}(s')(s) w(o_i|s', a) * \sum_{i=1}^{|O|} \sum_{s' \in S} g^{y_i}(s')(s) w(o_i|s', a) \right| \\
&= \left| \sum_{s' \in S} \sum_{i=1}^{|O|} \left(g^{x^<_i}(s')(s) * g^{y_i}(s')(s) \right) w(o_i|s', a) P(s'|s, a) \right| \\
&\leq \sum_{s' \in S} \sum_{i=1}^{|O|} \left| g^{x^<_i}(s')(s) * g^{y_i}(s')(s) \right| w(o_i|s', a) P(s'|s, a) \\
&\leq \sum_{s' \in S} \sum_{i=1}^{|O|} m_v w(o_i|s', a) P(s'|s, a) = m_v.
\end{aligned}$$

So far, we have shown that $\alpha^{a}_{N*2}(s) * \alpha^{a}_{N*2}(s) \le m_v$ As given a fixed action a. Now let $a^{<}_t$ and a_t be the optimal action that is chosen at time t in our exact solution and approximate solution, respectively.

Case I: $a^{<}_t = a_t$ At $t \in \{N*3, \ldots, 1\}$: because $a^{<}_{N*1} = a_{N*1}$, we have $\alpha^{<}_{N*2} * \alpha_{N*2} \le m_v$. Applying the same argument inductively, we can easily show that $\alpha^{<a}_t(s) * \alpha^{a}_t(s) \le m_v$ for all t, and this implies that

$$V^{<} * V = \sum_{s \in S} \alpha^{<}_1(s)\, b_0(s) * \sum_{s \in S} \alpha_1(s)\, b_0(s) \le \sum_{s \in S} \left(\alpha^{<}_1(s) * \alpha_1(s)\right) b_0(s) \le m_v,$$

where $b_0(s)$ is the initial belief at state s with $\sum_{s \in S} b_0(s) = 1$.

Case II: $a^{<}_t \neq a_t$ for some $t \in \{N*3, \ldots, 1\}$: recall that during the reduction process, our heuristic algorithm only removed some elements in A_{N*1} (since $l = 1$). This implies that the set of feasible actions at every time step $t \in \{N*3, \ldots, 1\}$ remains the same. Thus, the action $a^{<}_t$ is feasible at every time step t also for our heuristic algorithm. If the action $a^{<}_t$ is indeed picked by our heuristic algorithm, then $a_t = a^{<}_t$. By case I, we have our desired results, i.e., $V^{<} * V \le m_v$. If another action is chosen instead in our heuristic algorithm, i.e., $a_t \neq a^{<}_t$, then let $\overline{V}$ be the value of the policy where at each time t, we are forced to pick the same action $a^{<}_t$. Because V is the optimal solution in our heuristic algorithm, we must have $V \ge \overline{V}$. Using the proof from Case I, this implies that $V^{<} * V \le m_v$.

Induction step: Assume that $V^{<} * V \le m_v \cdot l$ is valid for $l = i$; show that the inequality holds when $l = i + 1$. From the base case, we observe that our worst case error bound holds constant if no additional reduction in the size of A_t for $t = 1, \ldots, N*2$ is performed. Thus without loss of generality, we can assume that the horizon of our POMDP problem has length 1. However, repeating the same argument as in the base case, we obtain that the worst case error increases by an additional factor of m_v.

To show that $V^{<} * V \le m_v \cdot V^{<} * V$, we first claim that when the observation o is always observable, $|A_t^{a,o}| = 1$ (we prove this below). When $|A_t^{a,o}| = 1$, there is no need to reduce the size of $A_t^{a,o}$. Thus, invoking $\text{reduce}(A_t^{a,o}, m_v, |D_p|)$ when o is observable does not produce additional error. Consequently, the error of $V^{<}$ is only induced by invoking $\text{reduce}(A_t^{a,o}, m_v, |D_p|)$ for these sets where o is partially observable. More specifically, recall that the construction of algorithm $\text{reduce}(A_t^{a,o}, m_v, |D_p|)$ implies that $\left|\alpha^{<}_{x_i}(a, o_i)(s) * \alpha_{y_i}(a, o_i)(s)\right| \le 2 d_v = \frac{m_v}{|O_p|}$ for all states s. Thus, indeed in the base case above, at time t, we have

$$\begin{aligned}
\left|\alpha^{<a}_{N*2}(s) * \alpha^{a}_{N*2}(s)\right| &= \left|\sum_{i=1}^{|D_p|} \alpha^{<}_{x_i}(a, o_i)(s) * \alpha_{y_i}(a, o_i)(s)\right| \\
&= \left|\sum_{i=1}^{|D_p|} \alpha^{<}_{x_i}(a, o_i)(s) * \alpha_{y_i}(a, o_i)(s)\right| \\
&\le \sum_{i=1}^{|D_p|} \left|\alpha^{<}_{x_i}(a, o_i)(s) * \alpha_{y_i}(a, o_i)(s)\right| \\
&\le |D_p| \cdot \frac{m_v}{|D_p|} = m_v.
\end{aligned}$$

Following the same induction above, we obtain the desired result.

To show that $\delta A_t^{a;o}\delta = 1$ when o is observable, we notice $w.o\delta\ddot{s}; a/$ is either 1 or 0 for $A_t^{a;o}$. In the case where $w.o\delta\ddot{s}; a/ = 0$ when $xing\ a; o$, then $A_t^{a;o} = \hat{0}`$; in the case where $w.o\delta\ddot{s}; a/ = 1$, we are simply back to the MDP case where the optimal value of the state is unique. □

2.16 Transition probability matrix

Recall the transition probability matrix from Ÿ 2.6.2:

$$P_t^a = \begin{array}{c} \text{from\_to} \\ \text{Dx} \\ \text{NC} \\ \text{C1} \\ \text{C2} \\ \text{Re} \\ \text{OD} \\ \text{Dt} \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Dx} & \text{NC} & \text{C1} & \text{C2} & \text{Re} & \text{OD} & \text{Dt} & \text{Abs} \end{array} \\ \begin{bmatrix} x_{Dx} & nc_{Dx} & Z_1 & c_{Dx}^2 & 0 & od_{Dx} & d_{Dx} & w \\ 0 & nc_{NC} & Z_2 & 0 & 0 & 0 & d_{NC} & w \\ 0 & nc_{C1} & Z_3 & c_{C1}^2 & e_{C1} & od_{C1} & d_{C1} & w \\ 0 & 0 & Z_4 & c_{C2}^2 & e_{C2} & od_{C2} & d_{C2} & w \\ x_{Re} & 0 & Z_5 & c_{Re}^2 & e_{Re} & od_{Re} & d_{Re} & w \\ 1 * d_{OD} & 0 & 0 & 0 & 0 & 0 & d_{OD} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

where $Z_1 = 1 * x_{Dx} * nc_{Dx} * c_{Dx}^2 * od_{Dx} * d_{Dx} * w; Z_2 = 1 * nc_{NC} * d_{NC} * w; Z_3 = 1 * nc_{C1} * c_{C1}^2 * e_{C1} * od_{C1} * d_{C1} * w; Z_4 = 1 * c_{C2}^2 * e_{C2} * od_{C2} * d_{C2} * w; Z_5 = 1 * x_{Re} * c_{Re}^2 * e_{Re} * od_{Re} * d_{Re} * w$:

De ne: Nod: the number of overdoses that maps non-negative integer inputs to either 0 or 1.

Nr: the number of relapses that maps non-negative integer inputs to 1, 2, 3, 4, 5, or 6.

Tu: time since last use of drugs, measured in days.

TA: treatment adherence takes value x 90%), 2(70%-90%), 3(%).

Let DL be the length of the detoxi cation program the patient chooses.

Variable	Action	Formula/Function of
x_{Dx}	action independent	$1 * \frac{1}{DL}$
nc_{Dx}	action independent	$.1 * x_{Dx} * od_{Dx} * d_{Dx} * w/ \bullet 0{:}1$
c_{Dx}^2	action independent	$.1 * x_{Dx} * od_{Dx} * d_{Dx} * w/ \bullet 0{:}7$
od_{Dx}	action independent	if Nod >0: 0.524/2-year (Wines et al. 2007) if Nod = 0: 0.11/2-year (Wines et al. 2007)
d_{Dx}	action independent	if Nr = 0 : 0:001DL (Krebs et al. 2017) if Nr > 0: 0.001

w	action independent	0.02/year (Morral et al. (1997), O'Toole et al. (2006), TERMORSHUIZEN2005231)
nc_{NC}	action dependent	-
d_{NC}	action independent	0.005/263-day
nc_{C1}	action dependent	-
c^2_{C1}	action dependent	-
$e_{C1}; e_{C2}$	action dependent	one year relapse rate of 72% is obtained from (Chalana et al. 2016) the e ect of treatment adherence is estimated from (Nosyk et al. 2009)
od_{C1}	action dependent	estimated from (Wines et al. 2007)
d_{C1}	action independent	0.008/263-day
c^2_{C2}	action dependent	-
od_{C2}	action dependent	estimated from (Wines et al. 2007)
d_{C2}	action independent	0.01/263-day
x_{Re}	NT	0.18/month
	M, B, IN, C	TA 1: 0.4/month;
	M, B, C	TA 2: 0.3/month; TA 3: 0.2/month
E	NT (Zarkin et al. 2005)	0.007/month
	M, B, IN, C (Zarkin et al. 2005)	TA 1: 0.018/month;
	M, B, C (Zarkin et al. 2005)	TA 2:0.016/month; TA 3: 0.0144/month
c^2_{Re}	NT (Zarkin et al. 2005)	0.004/month
	M, B, IN, C (Zarkin et al. 2005)	TA 1: 0.012/month;
	M, B, C (Zarkin et al. 2005)	TA 2:0.01/month; TA 3:0.0084/month
od_{Re}	action independent (Krebs et al. 2017)	0:001 • N od + 0:009
d_{Re}	action independent (Krebs et al. 2017)	0.021/263-day
d_{OD}	NT (Kelty and Hulse 2017), C	if T u x 29 + DL: 0.158 + 0:01 • N od
	M (Kelty and Hulse 2017)	if T u < 29 + DL: 0.25+0:01 • N od
	M (Kelty and Hulse 2017), IN	if T u x 29 + DL: 0.11+0:01 • N od
	B (Kelty and Hulse 2017), IN	if T u < 29 + DL : 0:15 + 0:01 • N od
	B (Kelty and Hulse 2017)	if T u x 29 + DL : 0:09 + 0:01 • N od

Table 2.6: The description of the parameters in the transition probability matrix. -: assumptions that we made to ll out the transition probability matrices.

2.17 Additional Figures

Figure 2.7: Additional Figures for Scenario 2 where the patient reacts to treatment B better than M but with a di erent magnitude than the average treatment dynamics. Top row: budget 15K. Bottom row: budget 21K. Left column: medium TA. Middle column: high TA. Right column: low TA. The optimal 's, from top left to bottom right are approximately [0.294, 0.234, 0.201, 0.185, 0.162, 0.154, 0.158], [0.534, 0.421, 0.429], [0.126, 0.058, 0.058], [0.300, 0.272, 0.256, 0.239 0.214, 0.202, 0.216], [0.551, 0.481, 0.483], and [0.146, 0.061, 0.061]; The expected QALDs received from treatments (excluding the terminal reward) are [284.91, 293.96, 294.88, 296.21, 296.32, 296.14, 298.78], [313.46, 315.26, 315.84], [229.57, 225.76, 237.29], [312.11, 318.27, 318.70, 318.72, 318.72 318.80, 321.43], [320.14, 322.28, 322.86], and [297.28, 304.21, 314.31].

Figure 2.8: Reduced urine tests the frequency of urine tests in Cases 2a 2d is reduced to one third of the original frequency, and we remove urine tests in Case 3. In all sub gures, the patient has medium TA. Top row: Scenario 1. Top left: 9K budget. Top right: 15K budget. Middle left: 21K budget, Scenario 1. Middle right: 9K budget, Scenario 2. Bottom left: 15K budget, Scenario 2. Bottom right: 21K budget, Scenario 2. The optimal, from top left to bottom right, are approximately [0.052, 0.065, 0.054, 0.045, 0.038, 0.037, 0.040], [0.294, 0.187, 0.139, 0.125, 0.110, 0.103 0.103], [0.300, 0.222, 0.181, 0.165, 0.146, 0.138, 0.141], [0.052, 0.039, 0.036, 0.033, 0.026, 0.029, 0.029 [0.294, 0.234, 0.201, 0.185, 0.162, 0.158, 0.158], and [0.300, 0.274, 0.256, 0.239, 0.214, 0.208, 0.216].

Figure 2.9: Additional gures for sensitivity analysis. Left: 9K budget, Scenario 1. Right: 15K budget, Scenario 2. the frequency of urine tests in Cases 2a 2d is reduced to one third of the original frequency, and we remove urine tests in Case 3. In both sub gures, the patient has low TA. Left: generate with a di erent matrix perturbation magnitude than that in Fig. 2.4(top right); right: generate with the same matrix perturbation as in Fig. 2.8. The optimal from left to right, are approximately [121.27, 130.31, 132.77, 129.19, 113.58, 132.38, 139.45] and [0.126, 0.081, 0.069, 0.065, 0.060, 0.058, 0.059]; the expected QALDs received from treatments (excluding the terminal reward) are [0.099, 0.065, 0.061, 0.061, 0.060, 0.056, 0.056] and [229.57, 232.98, 229.99, 212.17, 216.07 229.55, 237.29].

2.18 Case Extension A POMDP Formulation for Survey Devices

In this case, in addition to urine test results, we now have access to daily self-reports on cravings. However, a challenge arises because patients may not tell the truth all the time. In particular, a patient may provide true, noisy, or falsi ed information. We assume that monetary incentives are provided to encourage patients to respond to the surveys over 80% of the time (Serre et al. 2012).[19] We have the same setup for treatment dynamics, and belief update, as in Case 1, but a di erent observation matrix . Notice that if a patient tells the truth all the time, the case will become equivalent to Case 4; if a patient provides only noisy information, this case is equivalent to Case 1. Consequently, we are interested only in the cases where patients provide falsi ed information or change their behaviors throughout the program.

To classify a patient's behavior, we rst map the numbers 1 to 4 to the partially observable

[19] Any missing entries in the self-report of a patient can be regarded as either noisy or falsi ed information as needed in our case.

states by a function n: $n(NC) = 1$, $n(C1) = 2$, $n(C2) = 3$, and $n(Re) = 4$. We then find the worst health state (s_w) that a patient reported up to three days before the urine test, and compare it with the test result. If s_w matches the urine test, we say the patient is telling the truth; if $s_w = Re$, but we have a negative urine test result, we say the patient is providing noisy information; if $s_w \neq Re$, but the urine test result is positive, we say the patient is providing falsified information.[20] In particular, we define the distance $d(s_w, Re) = 4 * n(s_w)$. For example, if $s_w = NC$, then $d(NC, Re) = 4 * 1 = 3$. We also maintain an average distance and map the patient self-report state to the two closest states[21] accordingly: if the input label is NC and the average distance is 2.3, then we will map NC to C2 with a probability of 0.7 and to Re with a probability of 0.3.

We use the tuple (l_t, l_r, l_s) to keep track of the likelihood that a patient is telling the truth, providing noisy reports, or giving falsified information, where $l_t + l_r + l_s = 1$. After plugging in our estimated transition probability matrix in Case 1, $* nc_{ut+} > 0.99$ with nc_{ut+} defined in Case 1, so we assume that Re is fully observable in this case. The set of observations thus becomes patients' self-reports, and the observation matrix is $W = l_t \bullet W_T + l_r \bullet W_R + l_s \bullet W_S$, where $W_T = I_{8\bullet 8}$ is the observation matrix when a patient is completely truthful. W_R and W_S are the observation matrices when a patient is providing random information and strategic information, respectively, and they are defined as follows:

$$W_R = \begin{array}{c} \\ Dx \\ NC \\ C1 \\ C2 \\ Re \\ OD \\ Dt \\ Abs \end{array} \begin{array}{c} \begin{array}{cccccccc} s\_o & & & & & & & \end{array} \\ \begin{bmatrix} Dx & NC & C1 & C2 & Re & OD & Dt & Abs \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Let d be the average distance, $\hat{d}$ be the integer part of d, and $\grave{d} = d * \hat{d}$ be the fractional part

[20] It could also be the case that the patient is providing noisy information. However, this scenario will be captured by parameter l_t below.

[21] If the average distance is an integer, we map it to only one label.

of d. Then

$$W_S(d=1) = \begin{array}{c} \\ \text{Dx} \\ \text{NC} \\ \text{C1} \\ \text{C2} \\ \text{Re} \\ \text{OD} \\ \text{Dt} \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Dx} & \text{NC} & \text{C1} & \text{C2} & \text{Re} & \text{OD} & \text{Dt} & \text{Abs} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{d} & 1 * \hat{d} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Similarly, we can enumerate the cases when $d = 0, 2, 3$:

$$W_S(d=0) = \begin{array}{c} \\ \text{Dx} \\ \text{NC} \\ \text{C1} \\ \text{C2} \\ \text{Re} \\ \text{OD} \\ \text{Dt} \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Dx} & \text{NC} & \text{C1} & \text{C2} & \text{Re} & \text{OD} & \text{Dt} & \text{Abs} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \hat{d} & 1 * \hat{d} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \hat{d} & 1 * \hat{d} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$W_S(d=2) = W_S(d=3) = \begin{array}{c} \\ \text{Dx} \\ \text{NC} \\ \text{C1} \\ \text{C2} \\ \text{Re} \\ \text{OD} \\ \text{Dt} \\ \text{Abs} \end{array} \begin{array}{c} \begin{array}{cccccccc} \text{Dx} & \text{NC} & \text{C1} & \text{C2} & \text{Re} & \text{OD} & \text{Dt} & \text{Abs} \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

To perform the belief update, we first calculate the observation matrix $W_X = l_t \bullet W_T + l_r \bullet W_R + l_s \bullet W_S$, and then follow the same procedures as in Case 1.

Chapter 3

Causal Inference with Selectively Deconfounded Data

3.1 Introduction

Say that we wished to determine whether sleep deprivation affects the likelihood of developing Alzheimer's disease. While we might easily observe that across a large population, sleep deprivation is highly correlated with Alzheimer's, this fact alone is not sufficient to establish whether sleep deprivation actually causes Alzheimer's. This is because the two variables might share a common cause (a confounder) accounting for the observed association, e.g., a genetic mutation that causes both sleep disorders and Alzheimer's.

The field of causal inference is concerned with precisely this type of problem. Generally, we wish to estimate the effect of assigning a given treatment (e.g., sleep deprivation) on a given outcome (e.g., Alzheimer's disease) in the presence of possible confounders (e.g., a genetic mutation).[1] One of the most fundamental problems in causal inference is to estimate the average treatment effect (ATE), i.e., the difference in the average outcomes that would be observed if everyone in the population did (versus did not) receive the treatment. The challenge of estimating the ATE in spite of unobserved confounding has been formalized mathematically (Pearl 1995, Rubin 1974), making clear that, in general, no amount of observational data is sufficient to identify the ATE. Specifically, there can (and typically do) exist multiple models consistent with the joint distribution over the observed variables that suggest different values of the ATE.

To bypass the challenge posed by confounding, research in causal inference has produced a variety of methods that tend to fall into two broad groups. The first group consists of performing

[1] Note here that in the technical sense, any variable upon which we could conceivably intervene and thereby influence the outcome can be called a treatment.

experiments on members of the population of interest, typically by assigning treatments randomly. The advantage of such randomized controlled trials (RCTs) is that they can eliminate unobserved confounding. However, while RCTs are often regarded as the gold standard for estimating causal effects, active experimentation is often infeasible, e.g., for ethical or financial reasons. The second group consists of leveraging domain knowledge to impose structural assumptions that render the parameter of interest identifiable. For example, if we assume that a sufficient set of confounders are observed (e.g., satisfying the backdoor criterion (Pearl 1995)), then under some mild statistical assumptions, the ATE can be estimated from observational data. Alternatively, even with unobserved confounding, other structural assumptions (e.g., valid instrumental variables or mediators) can render a causal effect estimable from observational data.

All of this is to say that the academic literature typically addresses the setting in which the confounders are either observed for all samples or for none. However, in many applications, it may be possible to revisit individuals represented in the dataset to retrospectively observe additional variables, such as the values of postulated confounders. For example, genetic information is stable over time, and thus when a genetic mutation is suspected to be a confounder, we might contact members of a study retrospectively to perform an additional test. Moreover, note that, as with this genetic example, even when the revelation of a variable is feasible it may nevertheless be expensive. In such settings, while it might be prohibitively expensive to reveal a confounder's value for every sample, we might still hope to reveal its value for a selected subset of our data. Consider the following examples:

1. Drug Repositioning: Here, drugs that have already been approved for a given disease are applied to treat a separate disease. Since these drugs are already proven to be safe, repositioning can be done in half the time and for a tenth of the cost when compared with the typical drug discovery process. To generate candidates for repositioning, correlations between drugs and diseases can easily be mined from health records. However, these relationships could plausibly be confounded by other drugs, which could both (a) affect the target disease and (b) trigger a side effect causing the candidate drug to be taken. While existing longitudinal studies aimed at assessing a drug's efficacy for a given disease A might have taken care to collect data on suspected confounders that might influence both treatment and disease A, when evaluating the drug as a candidate to treat disease B, we might need to reveal the value of other confounders. Moreover, due to the high costs of contacting patients and performing additional tests, such data collection efforts would be costly. Even for other known confounders, due to the digital divide of health records across disconnected systems, an expensive, manual effort, might be needed to collect the required data.

2. Genetic Factors for Disease: In the process of establishing the cause of different diseases,

genetic mutations are often implicated as potential confounders, such as in our initial motivating example for Alzheimer's. Due to the cost of DNA sequencing, it might only be possible to observe the genetic confounder for a subset of patients.

Succinctly, this paper addresses ATE estimation with selectively deconfounded data. We assume that we are given a large set of confounded data. At the outset only the treatment and outcomes are observed, but we have the ability to deconfound any of these samples, i.e., to reveal the sample's confounder. Naively, one could deconfound an arbitrary set of samples, and estimate the ATE with standard methods using only the subset of data that was deconfounded. However, this would discard a substantial amount of data. Thus motivated, we ask the following questions:

1. What is the value of confounded data? Specifically, how much can we improve ATE estimation by incorporating confounded data, relative to approaches that rely on deconfounded data alone?

2. What is the additional value of selective deconfounding? Specifically, how much can we further improve ATE estimation by intelligently selecting which confounded data to deconfound based upon the (observed) values of the treatments and outcomes?

To our knowledge, this is the first paper that focuses on the case where ample (cheaply-acquired) confounded data is available, and we may select only a limited number of samples to (at considerable cost) deconfound.[2]

3.1.1 Our Contributions

We address these questions for a standard confounding graph where the treatment and outcome are binary, and the confounder is categorical. More generally, we introduce a class of optimization problems that we dub *selective deconfounding*, where the values of the confounders are initially unobserved and can be subsequently revealed. Concretely, our ultimate goal is to find *sample-efficient selection policies*—policies for deciding which samples to deconfound in order to estimate the ATE most accurately. Our contributions are threefold:

1. The Value of Confounded Data (Ÿ 3.3): We show (in Theorem 5) that a simple method for incorporating confounded data achieves a constant-factor improvement in ATE estimation error over using deconfounded data alone. Loosely, the reason for this is the following: in our non-parametric causal inference model, estimating the ATE boils down to estimating the parameters of

[2] Throughout this paper, we implicitly assume that the confounded data was sampled i.i.d. from the target population of interest (but that our policy for selecting data to deconfound need not be).

the data-generating distribution. The inclusion of (in nite) confounded data reduces the number of free parameters to be estimated, improving our estimates of the remaining parameters. Moreover, since the causal functional (the expression for the ATE) is non-linear in the parameters to be estimated, the error the estimated ATE can be much greater than the individual errors in parameter estimates. Thus, our improvements in parameter estimates yield greater bene ts in estimating treatment e ects. For binary confounders, our numerical results show that on average, over problem instances selected uniformly on the parameter simplex, our method achieves roughly a 2:5 factor reduction in ATE estimation error.

2. The Additional Value of Selective Deconfounding (Ÿ 3.4): We show that selective deconfounding can in fact reduce ATE estimation error further, and we propose our own policy for doing so. We rst establish (in Proposition 8) that there cannot exist an optimal policy for selective deconfounding, in the sense that no policy is universally optimal for all data-generating distributions. Thus, instead we compare our proposed policy against two benchmark policies (which we call A and B just for now), along di erent metrics. Assuming access to in nite confounded data, we nd that:

1. With respect to the upper bounds on each policy's sample complexity that we show (Theorems 6 and 7), the guarantee for our proposed policy (a) entirely dominates benchmark policy A, and (b) is independent of (and therefore robust to) the observed distribution over treatment and outcome.

2. With respect to the (upper bound on) sample complexity of each policy under its worst-case data-generating distribution (Corollary 3.4), our proposed policy (a) dominates both benchmark policies, and (b) is independent (and again robust to) the entire data-generating distribution.

3. Among all estimators, we show (in Theorem 8) that our proposed policy requires no more than twice as many samples as benchmark policy B in the worst case, whereas benchmark policy B may require in nitely more samples as our policy.

We extend our work to the scenario where only a nite amount of confounded data is present, demonstrating that our qualitative insights continue to apply.

3. Experimental Evidence (Ÿ 3.5): Our synthetic experiments suggest that our proposed policy dominates both benchmark policies when averaging over the unknown data-generating distributions. Moreover, our experiments also characterize those data-generating distributions most favorable/unfavorable for our policy. We validate our methods on COSMIC (Tate et al. 2019,

Cosmic 2019), a real-world dataset containing cancer types, genetic mutations, and other patient features, showing that the practical bene ts of our proposed sampling policy. We show that on average, to achieve an ATE estimation error of 0.006, our proposed selection policy reduces the number of deconfounded samples by a factor of up to 10 when compared with the two benchmark policies.

3.1.2 Related Work

Causal Inference Without Unobserved Confounders: Causal inference has been studied thoroughly under the ignorability assumption, i.e., no unobserved confounding (Neyman 1923, Rubin 1974, Holland 1986). Some approaches for estimating the ATE under ignorability include the backdoor adjustment (Pearl 1995, Huang and Valtorta 2006), using either outcome regression (Rubin 1974), inverse propensity score weighting (Rosenbaum and Rubin 1983, Hirano et al. 2003, McCa rey et al. 2004), matching (Dehejia and Wahba 2002), or the use of instrumental variables when causal structural models are assumed (Sargan 1958, Angrist et al. 1996). Some related papers look to combine various sources of information, for instance from RCTs and observational data to estimate the ATE (Stuart et al. 2011, Hartman et al. 2015, Rosenman et al. 2018). Other papers leverage machine learning techniques, such as random forests, for estimating causal e ects (Alaa and van der Schaar 2017, Wager and Athey 2018). Other techniques include using time-series data to estimate the ATE (Athey et al. 2016), and targeted learning (Van der Laan and Rose 2011). In the operations research and management science literature, several optimization problems have been proposed. Nikolaev et al. (2013) propose a new objective for matching such that the bias of the estimated treatment e ect is minimized. Under resource constraints, Gupta et al. (2020) propose a new method for selecting individuals for treatment intervention to maximize the worst-case aggregate intervention e ectiveness.

Under the ignorability assumption, missing data has also been thoroughly studied. Under the assumption that either the missing mechanism or the distribution of the complete data is correctly speci ed, doubly robust estimators (the estimators that remain consistent under this assumption) have been proposed when data is missing at random and when the missing probabilities are either known or can be parametrized (Robins et al. 1994, Hannah et al. 2010). Doubly-robust estimators have also been studied when outcome is missing not at random (Rotnitzky et al. 1998, Scharfstein et al. 1999), and been proposed for sequential (Robins 2000) and longitudinal (Bang and Robins 2005) missing data.

Causal Inference With Unobserved Confounders: Since an unaccounted for unobserved confounder can invalidate an estimate of the ATE, three lines of work attempt to address/remove

	Obs./exp. data	Confounders	Guarantee	Active
Robins et al. (1994)	Observational	Missing at random	Asymptotic	No
Kallus et al. (2018)	Both	Unconfounded exp. data	Asymptotic	No
This work	Observational	Selectively deconfounded	PAC	Yes

Table 3.1: Literature that are closely related to our paper

the ignorability assumption: one using observational data alone, another by combining confounded observational data with experimental (and thus unconfounded) data, and finally by conducting sensitivity analysis.

The first line includes papers using proxies (Miao et al. 2018) and mediators (Pearl 1995). Kuroki and Pearl (2014) identify graphical structures under which causal effect can be identified. Miao et al. (2018) propose to use two different types of proxies to recover causal effects with one unobserved confounder. Shi et al. (2018) extend the work by Miao et al. (2018) to multiple confounders. However, both methods require knowledge of proxy categories a priori and are not robust under misspecification of proxy categories. Louizos et al. (2017) use variational autoencoders to recover the causal effect under the model where when conditioned on the unobserved confounders, the proxies are independent of treatment and outcome. Pearl (1995) introduces the front-door adjustment, a procedure whereby the causal effect can be expressed as a functional that concerns only the (possibly confounded) treatment and outcome, and an (unconfounded) mediator that transmits the entire effect. This procedure was further improved by Tian and Pearl (2002) and Shpitser and Pearl (2006a,b).

The second line combines confounded observational and experimental data to estimate the ATE. Bareinboim and Pearl (2013) propose to combine observational and experimental data under distribution shift, learning the treatment effect from the experimental data and transporting it to the confounded observational data to obtain a bias-free estimator for the causal effect. Recently, Kallus et al. (2018) propose a two-step process to remove hidden confounding by incorporating experimental data, relaxing the assumption that the confounded data and experimental data have the same support region.

Finally, to better interpret the estimated ATE and to address the possibility of unobserved confounders, a third line of the work aims to test the robustness of the ATE estimate via sensitivity analysis. Note that this line of work does not address the existence of unobserved confounding, but rather examines how the estimated ATE would changed under one additional (hypothetical) confounder. These work include Cornfield et al. (1959), Rosenbaum (1987, 2002), Rosenbaum et al. (2010), Rosenbaum (2011, 2014), Shen et al. (2011), VanderWeele and Ding (2017), Zhao et al.

Figure 3.1: Causal graph with treatment T, outcome Y, and confounder Z

(2017), Miratrix et al. (2018). Fogarty and Small (2016) extend the analysis to multiple unobserved confounders by enforcing all unmeasured confounders to have the same impact on the treatment assignment probabilities.

Unlike most prior work, we (i) address confounded and deconfounded (but not experimental) observational data, (ii) perform nite sample analysis to quantify the relative bene t of additional confounded and deconfounded data towards improving our estimate of the average treatment e ect, and (iii) investigate sample-e cient policies for selective deconfounding.

3.2 Model and Estimator

We begin by introducing the process through which (we assume) the data is generated and observed (Fig. 3.1). Let T and Y be binary random variables representing the treatment and outcome, respectively. The binary values can be taken, for example, as indicating whether a particular treatment has been applied (in the case of T) and whether the outcome was successful (in the case of Y). We assume the existence of a (potential) confounder, denoted Z, that can take up to k categorical values. Although there only exists a single confounder in our model, because the variable is categorical, the model subsumes scenarios with multiple categorical confounders. We will expand on this assumption later on in this section.

Let $P_{Y;T;Z}$ denote the joint distribution of the random vector $Y; T; Z/$, the randomness corresponding to draws from a given population. We will use similar subscripts to denote conditional distributions (e.g., $P_{Y\ð T;Z}$) and marginal distributions (e.g., $P_{;Z}$). Confounded data then consists of i.i.d. draws from $P_{Y;T}$ (marginalized over the confounder Z), and deconfounded data consists of i.i.d. draws from $P_{Y;T;Z}$. Thus, the confounded and deconfounded data are $.y; t/$ and $.y; t; z/$ tuples, respectively. We use the term *selective deconfounding* to mean selecting a confounded data point $.y; t/$, and revealing the value of its confounder z; this corresponds to sampling z from $P_{Z\ðY;T}. ðy; t/$.

Our goal is to estimate the *average treatment e ect* (ATE). Now there are a number of (equivalent) formalizations of the ATE. One formalization follows the now-classical nomenclature established

by Pearl (2000): let the so-called do-conditional expression $P(Y = y \mid do(T = t))$ be defined as

$$P(Y = y \mid do(T = t)) := \sum_{z \in [k]} P_{Y \mid T,Z}(y \mid t, z) P_Z(z).$$

Intuitively, $P(Y = y \mid do(T = t))$ describes the probability of $Y = y$ when we force the treatment to take value t, regardless of what value it would have otherwise taken. For instance, in the Alzheimer's example, this corresponds to forcing an individual to sleep or not sleep.

The ATE then can be expressed, via the back-door adjustment (Pearl 1995), as

$$\begin{aligned} \mathrm{ATE}(P_{Y,T,Z}) &:= P(Y = 1 \mid do(T = 1)) - P(Y = 1 \mid do(T = 0)) && (3.1) \\ &= \sum_{z \in [k]} P_{Y \mid T,Z}(1 \mid 1, z) P_Z(z) - P_{Y \mid T,Z}(1 \mid 0, z) P_Z(z). && (3.2) \end{aligned}$$

In words, the ATE is the difference between the average outcome in the population were we to always administer the treatment, and the average outcome were we to never administer the treatment.

We make the following assumptions, all of which are by now standard in the causal inference literature.[3]

Assumption 1. The following hold:

1. Ignorability [Rosenbaum and Rubin 1983]: $\{Y_t\} \perp T \mid Z$, where Y_t is the outcome that Y would take if the treatment were t.

2. Consistency [Robins 1986]: The observed outcome for an individual with treatment t, is the same as the one that we would observe if we were to assign treatment t to individual i.

3. Positivity [Cole and Hernán 2008]: $P_{T \mid Z}(t \mid z) < 1$ for all z such that $P_Z(z) > 0$.

Recall our motivation we cannot calculate the ATE from confounded data alone.[4] Precisely, even given an infinite amount of confounded data (which corresponds to knowing $P_{Y,T}$ exactly), there can exist multiple distributions $P_{Y,T,Z}$ consistent with the confounded data, with different values for $\mathrm{ATE}(P_{Y,T,Z})$. So given confounded data alone, the only available options are to collect new deconfounded data, or selectively deconfound existing data.

[3] These assumptions are minimal: the ignorability assumption ensures the identifiability of the causal effect. The consistency rule is so ubiquitous in causal inference as to be axiomatic at this point (Pearl 2010). Finally, positivity ensures the estimability of the causal effect.

[4] We could if we intervene or make further assumptions on the structure of the causal graph, but these options are often unavailable, as in our motivating applications.

Aside: Model Generalizability: Our next step, in the coming subsection, will be to define the estimator for the ATE that we will study. But before that, we pause in this (optional) aside to address some of the components of our model that may appear to be restrictive at first glance. First, on the use of discrete variables: there do indeed exist applications and a rich literature, in which the variables are instead continuous. Now such variables can of course be quantized to fit our model, but more importantly, we note that the use of categorical (even binary) data is well-established in both theory (Bareinboim and Pearl 2013) and application (Knudson 2001, Rayner et al. 2016), and not merely a simplifying proxy for continuous data. Moreover, models with continuous variables, by and large, require some additional structural assumptions (e.g., linearity, an alternate parameterization, or smoothness). Our model requires no such assumption.

A second ostensible restriction is that we have a single confounder, whereas applications may have multiple confounders. Our model in fact subsumes these scenarios. In particular, absent additional distributional assumptions, our model captures multiple unobserved confounders by simple concatenation without loss (since we impose no limit on the number of classes). Now, one could make additional assumptions (indeed, a high-dimensional setting might necessitate such assumptions) that could render alternative algorithms applicable. However, there exist many applications where (a) the confounder is of moderate dimension; and (b) a practitioner would be dubious of any additional assumption (Bates et al. 2020).

Now concatenating multiple confounders into a single confounder does implicitly require that the set of confounders is either never observed or entirely observed, but this is also without loss so long as the costs of revealing the confounders are equal (e.g., this is the case in the genetic example). Intuitively, because we do not impose any independence assumption on the set of confounders, revealing all confounders offers maximal information on the joint distribution of the confounders. We formalize this statement in Ÿ 3.7.1.

Lastly, we have not considered pretreatment covariates. While for simplicity we focus only on the setting where the confounder can be retroactively observed, we show in Ÿ 3.7.2 that our model can be applied straightforwardly to handle additional pre-treatment covariates.

3.2.1 An ATE Estimator

We are ultimately concerned with the "value" of different sets of data: first to contrast combined (confounded and deconfounded) data against deconfounded data alone, and second to contrast different sets of selectively deconfounded data. Now the "value" of any data here is measured with respect to the accuracy with which that data can be used to estimate the ATE. It is worth emphasizing here that we are not directly concerned with the question of how data is used to estimate the ATE — the design of such ATE estimators is an active area of research (Robins et al.

1994, Robins 2000, Bang and Robins 2005), which we view as orthogonal to the questions we seek to answer. Our approach will be to fix a particular ATE estimator to be applied on all sets of data. We will now define this estimator, which is conceptually and algorithmically simple, and corresponds to known, well-studied estimators (see § 3.2.2).

To ease notation slightly, let p^z_{yt} denote the probability of event $\hat{}Y = y; T = t; Z = z`$, let $a_{yt}$ denote the probability of event $\hat{}Y = y; T = t`$, and let q^z_{yt} denote the conditional probability of event $\hat{}Z = z$ given $Y = y; T = t`$:

$$p^z_{yt} := P_{Y;T;Z}(y; t; z); \quad a_{yt} := P_{Y;T}(y; t); \quad q^z_{yt} := P_{Z|Y;T}(z|y; t):$$

We will further compact the notation by letting p denote the vector of all $4k$ (recall that Y and T are binary-valued, and Z takes one of k categorical values) values p^z_{yt}, in arbitrary order, and similarly for a and q. As a sanity check, the ATE can be computed entirely from p, but the reason to call out the decomposition in a and q (it is a decomposition in the sense that $p^z_{yt} = a_{yt} q^z_{yt}$) is that it will highlight the different information contained in confounded data (a) versus deconfounded data (q). Now on to our estimator, which differs slightly based on the type of data being used:

1. Deconfounded data alone: Given only deconfounded data, we first obtain empirical estimates for each p^z_{yt} using the maximum likelihood estimator (MLE), which is simply the empirical frequency, and we denote the corresponding estimates $\hat{p}^z_{yt}$. We then obtain our estimated average treatment effect, $\hat{ATE}$, by plugging $\hat{p}^z_{yt}$ directly into the definition of ATE in Eq. (3.2). Specifically, let $ATE(p) : [0; 1]^{4k} \to [-1; 1]$ denote the value of the ATE under the distribution p:

$$ATE(p) = \sum_z \left(\frac{p^z_{11}}{p^z_{y1}} - \frac{p^z_{10}}{p^z_{y0}} \right) \sum_{y;t} p^z_{yt} \tag{3.3}$$

   (this is just a re-writing of Eq. (3.2)). Then our estimator under deconfounded data alone:

$$\hat{ATE} = ATE(\hat{p}): \tag{3.4}$$

   While this estimator is quite simple, in § 3.2.2 we show that absent additional causal structural assumptions, it corresponds to the well-studied doubly-robust estimation method.

2. Incorporating confounded data: To assess the value of confounded data, recall we have decomposed $P_{Y;T;Z}$ into two components: (i) the confounded distribution $P_{Y;T}$, or a; and (ii) the conditional distribution $P_{Z|Y;T}$, or q. The process of deconfounding reveals the value of Z for one (initially confounded) sample, and so we gain no additional information about the joint distribution $P_{Y;T}$. Thus, a is estimated entirely using the confounded data, and the deconfounded data can then be used exclusively to estimate the conditional distribution q:

(a) In nite samples of confounded data : Much of our analysis concerns the case where the amount of confounded data is e ectively in nite, so that the confounded distribution a is known exactly. Analogous to Eq. (3.3), we let $\mathrm{ATE}(a;q) : [0;1]^4 \bullet [0;1]^k \to [*1;1]$ denote the value of the ATE under the distribution decomposed as and q:

$$\mathrm{ATE}(a;q) = \sum_z \left(\frac{a_{11} q^z_{11}}{a_{y1} q^z_{y1}} * \frac{a_{10} q^z_{10}}{a_{y0} q^z_{y0}} \right) \sum_{y,t} a_{yt} q^z_{yt} \; : \tag{3.5}$$

We estimate each value q^z_{yt} from deconfounded data using the MLE estimator, and denote the estimates as $\hat{q}^z_{yt}$. We then calculate our estimated ATE by plugging the a_{yt}'s and $\hat{q}^z_{yt}$'s into Eq. (3.5):

$$\hat{\mathrm{ATE}} = \mathrm{ATE}(a;\hat{q}) : \tag{3.6}$$

(b) Finite samples of confounded data : Finally, in the case where we have a nite number of samples of confounded data, we use the confounded data to produce estimates $\hat{a}_{yt}$, via the MLE, and calculate our estimated ATE by plugging the $\hat{a}_{yt}$'s and $\hat{q}^z_{yt}$'s into Eq. (3.5):

$$\hat{\mathrm{ATE}} = \mathrm{ATE}(\hat{a};\hat{q}) : \tag{3.7}$$

3.2.2 Aside: Our Estimator as the Doubly Robust Estimator

In this (optional) subsection, we show that when incorporating deconfounded data only, without making any additional assumptions about the causal structure, our estimator, Eq. (3.4), is the same as the one obtained by applying the well-established doubly-robust estimation method, i.e. there is no bene t of doubly robustness under our problem setup. We do so by showing our estimator is the same 1) as the one obtained through a generic outcome regression model, and 2) as the one obtained via the inverse-propensity score weighting (IPW) method. In addition, we note that under our current problem setup, the IPW estimator always yield the same estimators as the ones in our paper, Eqs. (3.4) - (3.7). Finally, in Ÿ 3.8, we show that with in nite confounded data, a straightforward extension of the outcome regression model leads to an optimization problem that is not well-de ned. This motivates us to provide estimator-independent theoretical guarantees in Ÿ 3.3.

Our Estimator as the Outcome Regression Estimator First, under deconfounded data only, we observe only the tuples $(y;t;z)$ (and we do not know the confounded distribution $P_{Y;T}$). Without making any additional causal structural assumptions, we can only estimate the conditional average

treatment effect (CATE) the treatment effect when conditioned on the value of the confounder Z by one-hot encoding the (t, z) tuples. We can then obtain the ATE by reweighting the CATE by $P(Z)$ (which can be estimated separately). We formally state the outcome regression process below.

We first note that the ATE when conditioned on $Z = z$ equals to $P_{Y|do(T,Z)}(1|1, z) * P_{Y|do(T,Z)}(1|0, z)$. Let u_{tz} be a binary random variable that takes the value 1 if $T = t$ and $Z = z$, and 0 otherwise. Let y_i^{tz} to denote the values of Y when $T = t$ and $Z = z$, where $i = 1, ..., N_{tz}$ and N_{tz} is the number of samples where $T = t$ and $Z = z$. Then, using the random variables u_{tz}'s, we can estimate the value of $P_{Y|do(T,Z)}(1|1, z)$ by running the following logistic regression $P_{Y|do(T,Z)}(1|1, z) = \sigma(w_{tz}u_{tz} + b_{tz})$, where w_{tz} and b_{tz} are the weight and bias respectively, and σ is the logistic function. Then, the MLE for $(w_{tz}u_{tz} + b_{tz})$ can be obtained through solving the following optimization problem:

$$\arg\max \sum_{i=1}^{N_{tz}} y_i^{tz} \log \sigma(w_{tz}u_{tz} + b_{tz}) + (1 * y_i^{tz}) \log(1 * \sigma(w_{tz}u_{tz} + b_{tz}));$$

and we obtain the MLE estimate $\hat{y}^{tz} = \frac{\sum_{i=1}^{N_{tz}} y_i^{tz}}{N_{tz}}$. Recall that p^z_{yt} is our estimated $P_{Y,T|Z}(y, t|z)$ (using the MLE estimator), which can be calculated by dividing the number of samples where $Y = y, T = t, Z = z$ by N_{tz}. Let N be the total number of deconfounded samples. Then,

$$\frac{\sum_{i=1}^{N_{tz}} y_i^{tz}}{N_{tz}} = \frac{N p^z_{1t}}{N p^z_{0t} + N p^z_{1t}} = \frac{p^z_{1t}}{p^z_{0t} + p^z_{1t}}.$$

Thus, the estimated CATE under the above logistic regression can be expressed as

$$P_{Y|do(T,Z)}(1|1, z) * P_{Y|do(T,Z)}(1|0, z) = \frac{\sum_{i=1}^{N_{1z}} y_i^{1z}}{N_{1z}} * \frac{\sum_{i=1}^{N_{0z}} y_i^{0z}}{N_{0z}} = \frac{p^z_{11}}{p^z_{01} + p^z_{11}} * \frac{p^z_{10}}{p^z_{00} + p^z_{10}}.$$

One could verify that the above expression is indeed the same as the estimated CATE that we would have obtained using Eq. (3.4). Finally, to obtain the ATE, we can estimate $P(z)$ separately using deconfounded data alone via the MLE and obtain $\hat{P}(z) = \sum_{y,t} p^z_{yt}$. Together, we conclude that outcome regression yields the same estimator as in Eq. (3.4).

Our Estimator as the IPW Estimator Next, we show that under deconfounded data alone, the IPW estimator yields the same estimator as in Eq. (3.4).

Let τ denote the estimated ATE. We define the propensity score $e(z) := P(T = 1|Z = z)$, and let $\hat{e}(z)$ be the estimated propensity score from the data. Let t_i, y_i, and z_i denote the value of the treatment, outcome, and confounder for the ith sample respectively. First, we recall that the IPW estimator takes the following form:

$$\tau = \frac{1}{N} \sum_{i=1}^{N} \frac{t_i y_i}{\hat{e}(z_i)} * \frac{1}{N} \sum_{i=1}^{N} \frac{(1 * t_i) y_i}{1 * \hat{e}(z_i)}. \tag{3.8}$$

Note that in our problem, Z takes a finite number of discrete values. Thus, using the notation above we can further decompose Eq. (3.8) as follows:

$$, = \frac{1}{N}\sum_z \left[\sum_{i=1}^{N_{1z}} \frac{y_i^{1z}}{e(z)}\right] * \frac{1}{N}\sum_z \left[\sum_{i=1}^{N_{0z}} \frac{y_i^{0z}}{1 * e(z)}\right] \tag{3.9}$$

$$= \sum_z \left(\frac{1}{e(z)} \frac{\sum_{i=1}^{N_{1z}} y_i^{1z}}{N}\right) * \sum_z \left(\frac{1}{1 * e(z)} \frac{\sum_{i=1}^{N_{0z}} y_i^{0z}}{N}\right) \tag{3.10}$$

$$= \sum_z \left(\frac{1}{e(z)} \frac{N p_{11}^z}{N}\right) * \sum_z \left(\frac{1}{1 * e(z)} \frac{N p_{10}^z}{N}\right) = \sum_z \left(\frac{1}{e(z)} p_{11}^z\right) * \sum_z \left(\frac{1}{1 * e(z)} p_{10}^z\right); \tag{3.11}$$

where the second to the last equality is due to the fact that $\sum_{i=1}^{N_{tz}} y_i^{tz} = N p_{1t}^z$. Now, to estimate $e(z) = P(t = 1 | z)$, we note that without making any additional causal structural assumptions, $e(z)$ can be expressed using our estimators as follows: $e(z) = \frac{\sum_{yt} p_{y1}^z}{\sum_{yt} p_{yt}^z}$, and $1 * e(z) = P(t = 0 | z) = \frac{\sum_{yt} p_{y0}^z}{\sum_{yt} p_{yt}^z}$. Plugging in these values back into Eq. (3.11), we recover the Eq. (3.4) exactly. Finally, we observe that the IPW estimator can be formally stated as

$$\sum_z \left[\frac{P(Y, T, Z = 1, 1, z)}{P(T = 1 | z = 1, z)} * \sum_z \left[\frac{P(Y, T, Z = 1, 0, z)}{P(T = 0 | z = 0, z)}\right]\right]. \tag{3.12}$$

Thus, under our problem setup, without making any additional assumption, Eq. (3.12) always yields the same estimate as Pearl's backdoor adjustment formula Eq. (3.2).

3.3 The Value of Confounded Data

We are now prepared to answer our first question: how much can we improve ATE estimation by incorporating confounded data? To that end, we first analyze and compare the sample complexity of the MLE estimator described in § 3.2 for deconfounded data alone and for augmented with an infinite amount of confounded data, while holding everything else the same. Throughout the paper, we measure the sample complexity of an estimator $\hat{ATE}$ as the number of samples required for $\hat{ATE}$ to be (ϵ, δ)-close to the true ATE, where the notion of (ϵ, δ)-close is defined as follows:

Definition 2. An estimator $\hat{ATE}$ is said to be (ϵ, δ)-close to the true ATE if it satisfies

$$P(|\hat{ATE} * ATE| < \epsilon) > 1 * \delta.$$

Deconfounded Data Alone: We begin with the baseline approach of using only deconfounded data that has been sampled according to the deconfounded distribution $P(v, z)$. The following theorem identifies a quantity of samples, m_{base}, which is sufficient to estimate the ATE to within a desired level of accuracy under the estimation process described above.

Theorem 5. Using deconfounded data alone, the estimator $\widehat{\mathrm{ATE}}(\cdot;p)$ as defined in Equation (3.4) is $(\varepsilon;\delta)$-close if the number of deconfounded samples is at least

$$m_{\text{base}} := C \max_{t,z} \left(\mathbb{E}_y \left[p_{yt}^z \right] \right)^{-2} = C \max_{t,z} \frac{1}{P_{T,Z}(t,z)\varepsilon^2},$$

where $C := 12.5k^2 \ln(8k/\delta)/\varepsilon^2$.

The proof of Theorem 5 (in § 3.9.1) relies on an additive decomposition of the estimation error on ATE in terms of the estimation error on the p_{yt}^z's, along with concentration via Hoeffding's inequality.

Incorporating Infinite Confounded Data: Now consider the setup in which we have deconfounded data, along with an infinite amount of confounded data, i.e., the marginal distribution $P_{Y,T}$ is known exactly. Analogous to Theorem 5, Theorem 6 identifies a sufficient number of deconfounded samples m_{nsp}.[5]

Theorem 6. Incorporating (infinite) confounded data, the estimator $\widehat{\mathrm{ATE}}(a;q)$ is $(\varepsilon;\delta)$-close if the number of deconfounded samples is at least

$$m_{\text{nsp}} := C \max_{t,z} \frac{\left(\sum_y a_{yt} \right)^2}{\left(\sum_y a_{yt} q_{yt}^z \right)^2} = C \max_{t,z} \frac{P_T(t)}{P_{T,Z}(t,z)\varepsilon^2}, \tag{3.13}$$

where $C := 12.5k^2 \ln(8k/\delta)/\varepsilon^2$.

The proof of Theorem 6 is included in § 3.9.5. A few observations can be made from comparing Theorems 5 and 6:

1. m_{nsp} is less than m_{base} for any underlying distribution $P_{Y,T,Z}$, highlighting the value of confounded data. In fact, the ratio $m_{\text{base}}/m_{\text{nsp}}$, in addition to being strictly greater than 1, can be arbitrarily large.

2. With respect to the accuracy parameters ε and δ, both m_{base} and m_{nsp} scale as $\varepsilon^{-2}\ln(\delta^{-1})$, irrespective of the underlying distribution $P_{Y,T,Z}$. Now it might be that a better scaling is achievable, either with a smarter estimator or a tighter analysis of our estimator, but the following minimax lower bound shows that this is not the case:

   Proposition 6. (Lower Bound with respect to ε and δ) Fix any confounded distribution and assume that infinite confounded data is given (or equivalently, $P_{t,y}$ is known). For any ATE

[5] The subscript nsp stands for "natural selection policy". The motivation for this name will be explained in the next section.

estimator, there exists an underlying distribution $P_{Y;T;Z}$, whose confounded distribution is $P_{Y;T}$, for which the number of deconfounded samples[6] required for the estimator to be ϵ, δ-close is at least $\Omega\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$.

The proof of Proposition 6 (§ 3.9.3) proceeds by construction. Proposition 6 states not only that Theorems 5 and 6 are tight with respect to ϵ and δ, but moreover the $\Omega\left(\frac{1}{\epsilon^2}\log\frac{1}{\delta}\right)$ sample-complexity is necessary even for selectively deconfounded data, as we will study in the next section.

3. Even fixing k, ϵ, and δ, both m_{base} and m_{nsp} can be arbitrarily large for certain underlying distributions $P_{Y;T;Z}$. In fact, even fixing any confounded distribution $P_{Y;T}$, both m_{base} and m_{nsp} can still be arbitrarily large for distributions $P_{Y;T;Z}$ consistent with $P_{Y;T}$. As in the previous point, this is necessary for any estimator:

   Proposition 7. (Lower Bound with respect to $P_{Y;T;Z}$) Fix any confounded distribution and assume that in finite confounded data is given (or equivalently, $P_{Y;T}$ is known). There exists $\epsilon, \delta > 0$ such that no ϵ, δ-close estimator exists. Specifically, for any number of deconfounded samples m,[7] there exist two underlying distributions $P^1_{Y;T;Z}$ and $P^2_{Y;T;Z}$ with the following properties:

   - Both of their confounded distributions are $P_{Y;T}$.
   - No algorithm can correctly identify both of them with probability more than $1-\delta$ using at most m deconfounded samples.
   - Their corresponding ATE's are ϵ apart: $\left|\mathrm{ATE}.P^1_{Y;T;Z}\right) - \mathrm{ATE}.P^2_{Y;T;Z}\right)\big| > \epsilon$.

   The proof of Proposition 7 (§ 3.9.2) is constructive, and relies on ϵ being chosen as a function of $P_{Y;T}$, and the distributions $P^1_{Y;T;Z}$ and $P^2_{Y;T;Z}$ being chosen as a function of ϵ and m. The construction relies on values of the conditional distribution $P_{Z|Y;T}$ approaching 0 and 1, and so later on, we will state certain guarantees with respect to a parameter γ such that all values of $P_{Z|Y;T}$ are bounded within an interval $[\gamma, 1-\gamma]$.

3.4 The Additional Value of Selective Deconfounding

Having established the potential value of confounded data, we turn now to the second question: what is the additional value of selective deconfounding? The path we will take to answering

[6] This applies when the deconfounded samples are generated according to $P_{Y;T;Z}$ as in this section, or are selectively deconfounded as in the following section.

[7] See footnote 6

this question is to analyze different policies for selective deconfounding, and compare their performances via instance-dependent sample complexity lower and upper bounds, and the instance-independent worst-case sample complexity upper bounds when incorporating an infinite amount of confounded data. Incidentally, this will yield a policy which we propose for use in practice. Lastly, we extend our analysis to the scenario where we have access only to a finite amount of confounded data.

Selective Deconfounding: One important consequence of our procedure for estimating the ATE in Ÿ 3.2 is that the four conditional distributions are estimated separately: the deconfounded data is partitioned into four groups, one for each $(y, t) \in \{0,1\}^2$, and then the quantities s^z_{yt} are estimated separately. This means that the procedure does not rely on the fact that the deconfounded data is drawn from the exact distribution $P_{Y,T,Z}$. In particular, the draws might as well have been made directly from the conditional distributions $P_{Z|Y,T}$.

Suppose now that we can draw directly from these conditional distributions $P_{Z|Y,T}$. This situation may arise when the confounder is fixed (like a genetic trait) and can be observed retrospectively. This leaves us with four sample selection options, namely selecting confounded samples from the four groups $(y, t) \in \{0,1\}^2$, to deconfound. The problem of selective deconfounding formally can be stated as: given a budget for selectively deconfounded samples, how should we allocate our samples among the four groups $(y, t) \in \{0,1\}^2$?

Selection Policies: A sample selection policy knows the confounded distribution (assuming infinite confounded data), and selects the number of samples to deconfound from each of the four groups $(y, t) \in \{0,1\}^2$. Equivalently, it is a mapping from confounded distributions a (recall that $a_{yt} = P_{Y,T}(y, t)$) to vectors $x := (x_{00}, x_{01}, x_{10}, x_{11})$, where x_{yt} indicates the proportion of samples allocated to each group. We will consider the following three sample selection policies:

1. Natural (NSP): $x_{yt} = a_{yt} = P_{Y,T}(y, t)$. This is *natural* in the sense that it is equivalent to drawing directly from $P_{Y,T,Z}$, as we did in the previous section. Since it can be implemented by deconfounding samples at random, it corresponds to *non-selective* deconfounding.

2. Uniform (USP): $x_{yt} = 1/4$. This splits the samples evenly across all four groups. While this policy does not depend on a, it requires selective deconfounding to implement.[8]

3. Outcome-weighted (OWSP): $x_{yt} = a_{yt} / (2 \cdot a_{yt}) = P_{Y|T}(y|t)/2$. This splits the samples evenly across treatment groups T (= 0 vs. 1), but within each treatment group, the number of samples is proportional to the outcome Y (= 0 vs. 1).

[8] We also observe that USP with 4X number of samples (we simply select data from all 4 categories to deconfound at each step) lower bounds the estimation error of the optimal selection policy in our problem.

While the particular form of OWSP appears to be the least intuitive, we will soon show that it provides an instance-independent guarantee when considering the worst-case conditional distributions q.

As a sanity check, note that Theorem 6 corresponds to the sample complexity upper bound for NSP.

3.4.1 Non-existence of an Optimal Policy

Before we analyze the sample complexities of the remaining sample selection policies, we first establish that there does not exist a sample selection policy that is optimal. Now the notion of optimality here needs to be defined carefully — for example, it should be independent of the particular choice of ATE estimator, and it should apply across a set of underlying distributions. To that end, we introduce the following definition:

Definition 3. Fix any $\varepsilon > 0$. For any sample selection policy π, any confounded distribution a, and any set of conditional distributions Q, define $m_\pi(a, Q)$ to be the minimum number of deconfounded samples such that there exists an estimator $\widehat{\text{ATE}}$ which achieves $P(|\widehat{\text{ATE}} - \text{ATE}(a)| < \varepsilon) > 1 - \delta$ for all $q \in Q$:

$$m_\pi(a, Q) = \min \left\{ m \in \mathbb{N} : \exists \widehat{\text{ATE}} \text{ s.t. } P(|\widehat{\text{ATE}} - \text{ATE}(a, q)| < \varepsilon) > 1 - \delta \; \forall q \in Q \right\}.$$

As another sanity check, note that Theorem 6 implies an upper bound on m_{nsp} for all a and Q, $m_{\text{nsp}}(a, Q) \le \sup_{q \in Q} m_{\text{nsp}}(a, q)$.

Given this definition, it is natural to call a policy π optimal for a given confounded distribution a and a given set of confounded distributions Q if (and only if) it achieves the minimum value $m_\pi(a, Q)$ across all policies. We can now formally state the non-existence of such an optimal policy for any a:

Proposition 8. For every confounded distribution a, there exists two sets of conditional distributions Q_1 and Q_2 such that any optimal sample selection policy under $m(a, Q_1)$ is not optimal under $m(a, Q_2)$.

The proof of Proposition 8 proceeds by construction (§ 3.9.4). Proposition 8 highlights the need to compare the proposed selection policies via various performance metrics.

3.4.2 Analysis of Policies

Returning to our three defined sampling policies, we will in this subsection show four results that each compare the policies along different metrics. Table 3.2 summarizes the comparison between the three selection policies.

	Lower Bound	Upper Bound	Worst-Case Bound	Independence
NSP	$\frac{C_1}{2}\max_{y,t}\frac{a_{yt}\cdot^3\sum_y a_{yt}/^2}{\cdot^3\sum_y a_{yt}q^z_{yt}/^2}$	$C\max_{t,z}\frac{^3\sum_y a_{yt}}{\cdot^3\sum_y a_{yt}q^z_{yt}/^2}$	$\frac{C}{2}\max_t\frac{1}{^3\sum_y a_{yt}}$	
USP	$\frac{4C_1}{2}\max_{y,t}\frac{a^2_{yt}\cdot^3\sum_y a_{yt}/^2}{\cdot^3\sum_y a_{yt}/^2}$	$C\max_{t,z}\frac{^3\sum_y 4a^2_{yt}}{\cdot^3\sum_y a_{yt}q^z_{yt}/^2}$	$\frac{4C}{2}\max_t\frac{^3\sum_y a^2_{yt}}{\cdot^3\sum_y a_{yt}/^2}$	
OWSP	$\frac{2C_1}{2}\max_{y,t}\frac{a_{yt}\cdot^3\sum_y a_{yt}/^2}{^3\sum_y a_{yt}}$	$2C\max_{t,z}\frac{\cdot^3\sum_y a_{yt}/^2}{\cdot^3\sum_y a_{yt}q^z_{yt}/^2}$	$\frac{2C}{2}$	$m_{owsp}\circ P_{Y,T}$ $M_{owsp}\circ P_{Y,T,Z}$

Table 3.2: Comparison between the instance-speci c lower bound, sample complexity upper bound (m), and the worst-case sample complexity upper bound (M). m_{owsp} dominates m_{usp}, but neither m_{owsp} nor m_{nsp} dominates the other. Instead, Theorem 8 establishes the worst-case relative performance of $_{owsp}$ and $_{nsp}$. The last column illustrates that with su cient samples, the sample complexity upper bound guarantee for OWSP is independent of the confounded distribution $P_{Y,T}$, and the worst-case sample complexity upper bound for OWSP is independent of the data generating distribution $P_{Y,T,Z}$.

1. Upper Bounds on Sample Complexity: First, we provide an upper bound of the sample complexity of $_{usp}.a;Q/$ and $_{owsp}.a;Q/$ for every a and Q by analyzing our algorithm (analogous to Theorems 5 and 6):

Theorem 7. Incorporating (in nite) confounded data, the estimate $\widehat{ATE}.a;q/$ is $.\ ;\ /$-close if the number of deconfounded samples, selected under the natural selection policy (NSP) is at least:

$$m_{nsp} := C\max_{t,z}\frac{^3\sum_y a_{yt}}{^3\sum_y a_{yt}q^z_{yt}{}^2} = C\max_{t,z}\frac{P_T.t/}{P_{T,Z}.t;z/^2}:$$

Under the uniform selection policy (USP):

$$m_{usp} := C\max_{t,z}\frac{4^3\sum_y a^2_{yt}}{^3\sum_y a_{yt}q^z_{yt}{}^2} = C\max_{t,z}\frac{4^3\sum_y P_{Y,T}.y;t/^2}{P_{T,Z}.t;z/^2}:$$

Under the outcome-weighted selection policy (OWSP):

$$m_{owsp} := C\max_{t,z}\frac{2\ ^3\sum_y a_{yt}\ ^2}{^3\sum_y a_{yt}q^z_{yt}{}^2} = C\max_{t,z}\frac{2}{P_{Z,T}.z,t/^2}:$$

Here, $C := 12{:}5k^2\ln.8k\_/\ ^{*2}$.

As a sanity check, note that the rst statement in Theorem 7 is a restatement of Theorem 6, reproduced here to make comparison with the rest of the theorem easier. The proof of Theorem 7 (Ÿ 3.9.5), which di ers from the proof of Theorem 5, requires a modi cation to Hoe ding's inequality, which we derive to bound the sample complexity of the weighted sum of two independent random variables. Theorem 7 points to two advantages of OWSP:

1. OWSP has the nice property that m_{owsp} does not depend on $P_{Y;T}$. This means that when we have sufficient data, the sample complexity upper bound guarantee of OWSP is consistent across all confounded distributions $P_{Y;T}$.

2. USP is entirely dominated by OWSP: $m_{usp} \geq m_{owsp}$, since $4a_{0t}^2 + 4a_{1t}^2 + 2 \cdot a_{0t} + a_{1t}/^2 = 2 \cdot a_{0t} + a_{1t}/^2 \geq 0$.

We might hope for a similar result by comparing m_{owsp} with m_{nsp}, but neither strictly dominates the other, and in fact Proposition 8 rules out the possibility of finding a policy which strictly dominates all others.

2. Worst-Case Upper Bounds: Taking a slightly different tack, we might consider computing what Theorem 7 guarantees for each policy, across all possible values of the conditional distribution q this is a reasonable re-interpretation of Theorem 7 since we do not know the value of q in advance. One problem with this approach is that for any confounded distribution, each of the three values m_{nsp}, m_{usp}, and m_{owsp}, can be made arbitrarily large (which is consistent with the lower bound in Proposition 7, which recall applies to any choice of estimator) by taking certain values of q to be close to 0 or 1. So instead of considering all possible values of q, we will parameterize this entire analysis by a constant $\ddot{E} \in .0; 1\_2/$ (Worst-Case Upper Bound) Fix $\ddot{E} \in .0; 1\_2/$ and let $Q := \{q : q_{yt}^z \ddot{E} [\ ; 1 - \ ]\ \text{Å}\ y; t; z\}$. Then,

$$M_{nsp} := \max_{q \ddot{E} Q} m_{nsp} = \frac{C}{2} \max_t \frac{1}{\sum_y a_{yt}};$$

$$M_{usp} := \max_{q \ddot{E} Q} m_{usp} = \frac{4C}{2} \max_t \frac{\sum_y a_{yt}^2}{.\sum_y a_{yt}/^2};$$

$$M_{owsp} := \max_{q \ddot{E} Q} m_{owsp} = \frac{2C}{2}:$$

A few observations to make on Corollary 3.4:

1. The maximum of m_{nsp}, m_{usp}, and m_{owsp} are always obtained at $q_{yt}^z = \ $ for some $t; z$. This further justifies for our choice of parameterization by $\ddot{E}$.

2. M_{owsp} is independent of $P_{Y;T;z}$. This means that when data is sufficient, the worst-case sample complexity upper bound for OWSP is consistent across all data generating distributions.

3. OWSP has the lowest worst-case bound:

$$M_{owsp} \leq M_{nsp} \quad \text{and} \quad M_{owsp} \leq M_{usp}:$$

To see the first inequality, note that $\min \max_t 1\_\sum_y a_{yt}/$ is achieved when $\sum_y a_{yt} = 1\_2$, so $\max_t 1\_\sum_y a_{yt}/ \geq 2$. To see the second inequality, note that for binary-valued Y, we have $2\sum_y a_{yt}^2 \geq \sum_y a_{yt}^2 + 2 \ _y a_{yt} = .\sum_y a_{yt}/^2$.

3. Estimator-Independent Upper Bounds: Zooming back out to our goal of comparing the three di erent selection policies, Theorem 7 e ectively eliminated USP since it is dominated by OWSP i(m_{usp} x m_{owsp}) for everydata-generating distribution. Regarding the comparison between NSP and OWSP, Corollary 3.4 showed that OWSP dominates NSP for every confounded distribution a, under each policy's worst-case instance cpTo expand on the comparison between OWSP and NSP, next we provide a stronger guarantee that holds arbitrary estimators. Recall from De nition 3 that $_x$.a; Q/ is the minimum number of deconfounded samples, collected under policy x, such that some estimator is (,)-close for all q Ë Q.

The following result (Proof in Ÿ 3.9.6) establishes that $_{nsp}$.a; Q/ may be arbitrarily larger than $_{owsp}$.a; Q/, but $_{owsp}$.a; Q/ is never more than twice as large:

Theorem 8. Fix any Ë .0; 1_2/ For any Ë .0; 0:5 * 2 .1 * /], there exist confounded distributions a, and Q Ï Q , such that $_{owsp}$.a; Q/_ $_{nsp}$.a; Q/ is arbitrarily close to zero. In addition, for all a and Q, $_{owsp}$.a; Q/ w 2 $_{nsp}$.a; Q/:

4. Lower Bounds on Sample Complexity: Finally, we show lower bounds on $_{nsp}$.a; Q /; $_{usp}$.a; Q /; and $_{owsp}$.a; Q / that are analogous to Theorem 6:

Theorem 9. (Lower Bound) Fix any Ë .0; 1_2/ and any a. Then,

$$_{nsp}.a; Q / \times \frac{C_1}{2} \max_{y;t} \frac{a_{yt} .^3 _{y^{\cdot\cdot}} a_{y^{\cdot\cdot}t},/^2}{.^3 _{y^{\cdot\cdot}} a_{y^{\cdot\cdot}t}/^2};$$

$$_{usp}.a; Q / \times \frac{C_1}{2} \max_{y;t} \frac{4a^2_{yt} .^3 _{y^{\cdot\cdot}} a_{y^{\cdot\cdot}t},/^2}{.^3 _{y^{\cdot\cdot}} a_{y^{\cdot\cdot}t}/^2};$$

$$_{owsp}.a; Q / \times \frac{C_1}{2} \max_{y;t} \frac{2a_{yt} .^3 _{y^{\cdot\cdot}} a_{y^{\cdot\cdot}t},/^2}{^3 _{y^{\cdot\cdot}} a_{y^{\cdot\cdot}t}};$$

where $t_{,}$ = 1 * t and C_1 × . k * 1/ ^{2} ln. ^{*1}/ ^{*2}.

The proof (Ÿ 3.9.7) proceeds by construction. When comparing the constants C and C_1, we observe that the upper and lower bounds match in ; and , demonstrating the relative tightness of our analysis.

3.4.3 Incorporating Finite Confounded Data

We have now shown that given an in nite amount of confounded data, OWSP outperforms the NSP in the worst case. However, in practice, the confounded data will be nite. Recall that in this case, deconfounding reveals the value of Z for one (initially confounded) sample, and thus we gain

no additional information about $P_{Y,T}$. Thus, these n confounded data provide us with an estimate of the confounded distribution $P_{Y,T}(y,t)$, a_{yt}. To check the robustness of OWSP, we extend our analysis to handle finite confounded data. With κ_{yt} defined as above, we can derive a theorem analogous to Theorems 5-7:

Theorem 10. Given n confounded and m deconfounded samples, with m, $\mathrm{ATE}_a(q)$ is (ϵ, δ)-close when

$$\min_{y,t,z} \frac{\left(a_{yt} q^z_{yt}\right)^2}{\frac{1}{\kappa_{yt} m} + \frac{(q^z_{yt})^2}{n}} = \min_{y,t,z} \left(\frac{P_{T,Z}(t,z)^2 a}{\sqrt{\frac{1}{\kappa_{yt} m} + \frac{(q^z_{yt})^2}{n}}} \right)^2 \geq 4C. \tag{3.14}$$

Here, $C := 12.5 k^2 \ln(8k/\delta)/\epsilon^2$.

The proof of Theorem 10 (§ 3.9.8) requires a bound we derive for the product of two independent random variables. A few results follow from Theorem 10. First, a quick calculation shows that when m is held constant, $P(|\widehat{\mathrm{ATE}}_a(q) - \mathrm{ATE}_a(q)| \geq \epsilon)$ remains positive as $n \to \infty$. This means that for a certain combinations of ϵ, δ, n, there does not necessarily exist a sufficiently large m s.t. $P(|\widehat{\mathrm{ATE}}_a(q) - \mathrm{ATE}_a(q)| \geq \epsilon) \leq \delta$ can be satisfied. However, when there exists such an m, then $m \leq \max_{y,t,z} \kappa_{yt}^{-1} P_{T,Z}(t,z)^{-2} (4C/\epsilon^2 - (q^z_{yt})^2/n)^{-1}$. Although Theorem 10 does not recover Theorem 7 exactly when $m \to \infty$,[9] it provides insights into the relative performance of our sampling policies. Moreover, a conclusion that is similar to Theorem 8 holds: $m_{owsp} \geq m_{nsp}$ w.p. 2, and there exist distributions $P_{Y,T,Z}$ such that $m_{owsp} - m_{nsp}$ is arbitrarily small. Theorem 10 also implies that when $n \geq (q^z_{yt})^2 \kappa_{yt} m \approx (y,t) \in \{0,1\}^2$, the majority of the estimation error comes from not deconfounding enough data. This is because when the number of confounded data that we have is more than $O(m)$, the error on the ATE in Eq. (3.14) is dominated by fact that we have enough deconfounded data. Put another way, for a given m, it is sufficient to have $n = O(m)$ confounded samples.

One new issue that arises with finite confounded data is that a sampling policy may not be feasible because there are not enough confounded samples to deconfound. This does not happen for NSP (assuming $m \leq n$), but can occur for USP and OWSP. When this happens, we approximate the target sampling policy as closely as is feasible (see § 3.5.2).

3.5 Experiments

Since the upper bounds that we derived in § 3.3 are not necessarily tight, we first perform synthetic experiments to assess the tightness of our bounds. For the purpose of illustration, we focus on

[9] We could apply Lemma 5 to obtain a bound that recovers Theorem 7 exactly when $\mathbb{E}[\cdot] \to 0$. However, this method does not give us sufficient insights into the comparative performance of our sampling policies.

binary confounders Z throughout this section, with $q_{yt} = P_{Z=1|Y,T}(y,t)$ and $q := (q_{00}, q_{01}, q_{10}, q_{11})$. We first compare the sampling policies in synthetic experiments on randomly chosen distributions $P_{Y,T,Z}$, measuring both the average and worst-case performance of each sampling policy. In terms of average performance, we find that OWSP outperforms NSP and USP (Fig. 3.2), and there exists data generating distributions in which OWSP underperforms NSP and USP (Fig. 3.3). When averaged over the conditional distribution, OWSP outperforms both NSP and OWSP (Fig. 3.4). In addition, we numerically investigate the data generating distributions in which OWSP outperforms NSP and USP (Fig. 3.4). We discover that 1) the advantage of OWSP over NSP is the largest when the treatment group is highly inbalanced, and 2) the advantage of OWSP over USP is the largest when the outcome is highly inbalanced within each treatment group. We then measure the effect of having finite (vs. infinite) confounded data (Fig. 3.5), demonstrating that OWSP is robust under finite confounded data. Finally, we test the performance of OWSP on real-world data taken from a genetic database, COSMIC, that includes genetic mutations of cancer patients (Tate et al. 2019, Cosmic 2019) (Fig. 3.6), showing the benefit of OWSP in real-world applications. Because this is (to our knowledge) the first paper to investigate the problem of selective deconfounding, the methods described in Ÿ 3.1.2 are not directly comparable to ours.

3.5.1 Infinite Confounded Data: Synthetic Experiments

Assuming access to infinite confounded data, we experimentally evaluate all four sampling methods for estimating the ATE: using deconfounded data alone, and using confounded data that has been selected according to NSP, USP, and OWSP. We evaluate the performance of four methods in terms of the absolute error, $|\widehat{\text{ATE}} - \text{ATE}|$. Because the variance of our estimators cannot be analyzed in closed form, we report the variance of the absolute error, averaged over different instances.

Performance on Randomly Generated Instances: We first evaluate the four methods over a randomly-selected set of distributions. Fig. 3.2 was generated by averaging over 13,000 instances, each with the distribution $P_{Y,T,Z}$ drawn uniformly from the unit 7-Simplex. Every instance consists of 100 replications, each with a random draw of 1,200 deconfounded samples. The absolute error is measured as a function of the number of deconfounded samples in steps of 100 samples. Fig. 3.2 (top left) compares the use of deconfounded data along with the incorporation of confounded data selected naturally (as in the comparison of Theorems 5 and 7). It shows that incorporating confounded data yields a significant improvement in estimation error. For example, achieving an absolute error of 0.02 using deconfounded data alone requires more than 1,200 samples on average, while by incorporating confounded data, only 300 samples are required. We observe that by incorporating infinite amount of confounded dataset, the variance of our estimator has

decreased dramatically. Having established the value of confounded data, Fig. 3.2 (top middle) compares the three selection policies. We find that, when averaged over joint distributions, OWSP outperforms both NSP and USP in terms of both the average absolute error and the variance. Fig. 3.2 (top right) compares the average squared error of the three selection policies. We find that OWSP outperforms NSP and USP in terms of estimation bias as well. To compare the performance of our sampling policies on an instance level, we provide three scatter plots in Fig. 3.2 (middle), each containing the 13,000 instances in the top figures and averaged over 100 replications. The number of deconfounded samples is fixed at 1,200. We observe that OWSP outperforms NSP and USP in the majority of instances. Furthermore, to compare the variance of different sampling policies on an instance level, we provide three additional scatter plots in Fig. 3.2 (bottom). At each level of deconfounded samples, all figures in the bottom row of Fig. 3.2 contain the 13,000 instance in the top figures. Each dot is calculated by taking difference between the variances of selected sampling policies, where the variance is calculated using the same 100 replications contained in the top figures. Fig. 3.2 (bottom left) contains the difference between the variance of NSP and the variance of USP on the same instances. A dot with a positive y-axis value represents that the variance of USP on a particular instance is lower than that of NSP at the given level of deconfounded data. We observe that the variance of USP is lower than NSP on the majority of instances, and vice versa if the y-axis value is negative. Similarly, a dot with a positive y-axis value on Fig. 3.2 (bottom middle and right) represents that the variance of OWSP is lower than these of NSP and USP, respectively, on a particular instance at a given level of deconfounded data. We observe that the variance of OWSP is lower than these of NSP and USP on the majority of instances.

Worst-Case Instances: In Fig. 3.3, we evaluate the performance of the three selection policies on joint distributions chosen adversarially against each. The three sub-figures (the columns) correspond to instances where NSP, USP, and OWSP perform the worst, respectively, from the left to the right. Each sub-figure is further subdivided: the top contains results for the single adversarial example while the bottom is averaged over 500 q's sampled uniformly from $[0, 1]^4$. The absolute error is averaged over 10,000 replications in the top figures and 500 in the bottom. In all cases, we draw 500 deconfounded samples and measure the absolute error in steps of 50 samples. Fig. 3.3 (left) validates Corollary 8. We observe that when the distribution is heavily skewed towards $Y = 0, T = 0$, OWSP and USP significantly outperform NSP. Fig. 3.3 (middle) shows that USP can underperform NSP, but when averaged over all possible values of q, USP performs better than NSP. Fig. 3.3 (right) illustrates that OWSP can underperform NSP and USP, but, when compared with the left and middle column, the performance of OWSP is close to that of NSP and USP. In Fig. 3.3 (bottom), when averaged over all possible values of q, OWSP outperforms

both. Finally, OWSP's variance is the lowest across all scenarios. Ÿ 3.10 provides representative examples in which each of these joint distributions could appear.

Insights: To better understand the properties of the confounded distributions in which OWSP performs better than its counterparts, we conduct additional experiments. Fig. 3.4 is generated with a di erent set of 13,000 data generating distributions than Fig. 3.2. In particular, the 13,000 distributions of $P_{Y,T,Z}$ is generated with 130 di erent confounded distributions a, and under each confounded distribution, we generate 100 di erent conditional distributions q. Similar to Fig. 3.2, every instance consists of 100 replications, and the absolute error is measured as a function of the number of deconfounded in steps of 100 samples. Fig. 3.4 (top row) investigate the relationship between the performance of a pair of selected methods and the level of confoundedness in an instance. Speci cally, Fig. 3.4 (top row) contain 13,000 dots, each representing one instance and the number of deconfounded samples is xed at 1,200. A dot with a positive y-axis value in top left gure represents that in this particular instance, USP yields a smaller average absolute estimation error than NSP, and vice versa if the y-axis value is negative. Similarly, in the top middle and right gures, a dot with a positive y-axis value represents that OWSP yields a smaller average absolute estimation error than NSP and USP, respectively. The level of confoundedness (the y-axis) is the absolute di erence between the true ATE of an instance and the confounded ATE, i.e., $|ATE - (P_{Y|do(T=1)} - P_{Y|do(T=0)})|$, i.e., the level of confoundedness equals $|ATE_q * P_{Y|do(T=1)} + P_{Y|do(T=0)}|$. Fig. 3.4 (top row) demonstrate that there is no obvious association between the level of confoundedness and the relative performance of our algorithms.

Next, we investigate the association between the performance of a pair of selected methods and the level of treatment inbalance of an instance, where the latter is calculated by $|\gamma_{y1} * \gamma_{y0}|$. Fig. 3.4 (middle row left) shows that USP works better than NSP when averaged over the (initially unobserved) conditional distribution q on the majority of instances. In particular, the bene t of USP over NSP increases when the treatment inbalance of an instance increases. Fig. 3.4 (middle row middle and right) rst validate the observation that we made in Fig. 3.3 when averaged over the conditional distributions q, OWSP outperforms NSP and USP on almost all instances. In addition, Fig. 3.4 (middle row middle) shows that the treatment inbalance of an instance alone explains the scenarios in which OWSP signi cantly outperforms NSP, i.e., when the treatment is signi cantly inbalanced in the observed confounded data. However, Fig. 3.4 (middle row right) shows that the bene t of OWSP over USP cannot be explained using the level of treatment inbalance alone.

Finally, we investigate the relationship between the relative performance of a pair of selected methods and the level of outcome inbalance within each treatment group in Fig. 3.4 (bottom). In particular, the level of inbalance within each treatment group is calculated by taking the maximum outcome inbalance of each treatment group, i.e., $\max_t(|a_{1t} - \gamma_{yt} * 0.5|)$. Fig. 3.4 (bottom right)

illustrates that the bene t of OWSP over USP increases when the maximum level of outcome inbalance within each treatment group increases.

3.5.2 Finite Confounded Data

Approximate Sampling policies under nite confounded data: To deconfound according to NSP with nite confounded data is to deconfound the rst n confounded data. For USP, we split the samples to the four groups as evenly as possible. That is, we max out the bottleneck group/groups and distribute the excess data as evenly as possible among the remaining groups. Under OWSP, we have $x_{yt} = a_{yt} \_ ^3 {}_y a_{yt}$, and when implementing OWSP, we will rst ensure that the deconfounded samples are split as evenly as possible across treatment groups, and then within the each group, we split the samples close as possible according to the outcome ratio.

Results: Given only n confounded data, we test the performance of the OWSP against that of NSP and USP. In Fig. 3.5, the absolute error is measured as a function of the number of confounded samples in step sizes that increment in the log scale from 100 to 10,000 while xing the number of deconfounded samples to 100. Fig. 3.5 (left) is generated by averaging over 13,000 instances, each consisted of 100 replications, and it compares three o ine sampling selection policies. Since when we only have 100 confounded samples, the three sampling policies are identical, the error curves corresponding to NSP, USP and OWSP start at the same point on the top left corner. We observe that as the number of confounded samples increases, OWSP quickly outperforms NSP and USP on average, and the gaps between OWSP and the other two selection policies widen. Since we x the number of deconfounded samples to be 100, all three sampling policies are equivalent when there are only 100 confounded samples in the dataset (i.e., we need to deconfound all confounded samples in all cases), and the average absolute errors of the three selection policies do not converge to 0 in Fig. 3.5.

Fig. 3.5 (middle) contains the 13,000 instances described above averaged over 100 replications. It compares the performance of OWSP with that of the NSP on an instance level. Similarly, Fig. 3.5 (right) compares the performance of OWSP with that of the USP. In both gures, we x the number of confounded samples to be 681. We observe that OWSP dominates NSP and USP in the majority of instances by both the absolute error and variance. Note that if we x the number of confounded samples and increase the number of deconfounded samples (with n), we observe that OWSP dominates USP and NSP when the number of deconfounded samples are small, and the gap shrinks as the number of deconfounded samples increases. When $n = n$, all three methods are equivalent.

3.5.3 Real-World Experiments: Cancer Mutations

Data Previously, we chose the underlying distribution $P_{Y,T,Z}$ uniformly from the unit 7-Simplex. However, real-world problems of interest may not be uniformly distributed. Since causal-inference methods can be hard to validate as the true causal effect is almost never observed, to illustrate the practicality of our methods, we consider a real-world observational dataset, picking three variables to be the outcome, treatment, and confounder, and artificially hiding the confounder for some examples. Finally, we evaluate our proposed sampling methods under the assumption that we have access to infinitely many confounded samples. The Catalogue Of Somatic Mutations In Cancer (COSMIC) is a public database of DNA sequences of tumor samples. It consists of targeted gene-screening panels aggregated and manually curated over 25,000 peer reviewed papers. We focus on the variables *primary cancer site*, and *gene*. Specifically, for 1,350,015 cancer patients, we observe their type of cancer, and for a subset of genes, whether or not a mutation was observed in each gene.

Causal Models In our experiments, we designate cancer type as the outcome, a particular mutation as the treatment, and another mutation as the confounder. This setup seems reasonable because it is well known that multiple genetic mutations are correlated with individual cancer types (Knudson 2001), and that mutations can cause both cancer itself and other mutations. As a concrete example, mutations in the genes that code RNA polymerases (responsible for ensuring the accuracy of replicating RNA) are found to increase the likelihood of both other mutations and certain cancer types (Rayner et al. 2016). The setting where the treatment mutation and cancer outcome are observed and the confounding mutation is unobserved is plausible because it is common that the majority of patients only have a subset of genes sequenced (e.g., from a commercial panel). For the purpose of illustration, we assume there is no other unobserved confounders in this subsection.

The top 6 most commonly mutated genes were selected as treatment candidates. For each combination of a cancer type and one of these genes, we removed patients for whom this gene was not sequenced, and kept all pairs that had at least 40 patients in each of the four treatment-outcome groups (to ensure our deconfounding policies would have enough samples to deconfound). This procedure gave us 275 unique combinations of a cancer (outcome), mutation (treatment), and another mutation (confounder). Since on average, each {cancer, mutation, mutation} tuple contains around 25,883 patients, we took the estimated empirical distribution as the data-generating distribution and applied the ATE formula described in Ÿ 3.3 to obtain the true ATE. To model the unobserved confounder, we hid the values of the confounder, only revealing the value to a sampling policy when it requested a deconfounded sample. We compared the use of deconfounded

data along with the incorporation of confounded data under the three sampling selection polices: NSP, USP, and OWSP.

Results: Fig. 3.6 (left) was generated with the 275 instances each repeated for 10,000 replications. The absolute error is measured as a function of the number of deconfounded samples in step sizes of 15. First, similar to Fig. 3.2, we observe that incorporating confounded data reduces both the absolute estimation error and the variance of the estimator by a large margin. Note the improvement of OWSP over NSP is larger in this case as compared to that seen in Fig. 3.2. Furthermore, when the number of deconfounded samples is small, OWSP outperforms USP. Note that Fig. 3.6 (left) does not start with 0 because absent any deconfounded data, the estimated ATE is the same for all sampling policies. In Fig. 3.6 (middle, right), we fix the number of deconfounded samples to be 45 and compare the performance of OWSP against that of NSP and USP, respectively. Both figures contain the 275 instances in the left figure, averaged over 10,000 replications. We observe that under this setup, OWSP dominates NSP in all instances, and outperforms USP in the majority of instances.

3.6 Conclusion

Although extensive studies have been conducted in causal inference, none addresses the case where revealing the value of the confounder is the only option to estimate the causal effect. In this paper, we propose the problem of causal inference with selectively deconfounded data, and provide a set of non-adaptive sample selection policies. Our theoretical results upper bound the amount of deconfounded data required under each sample selection policy and provide insights for why the outcome-weighted selection policy works better on average than natural selection policy. Furthermore, we conduct extensive experiments using both synthetic and real-world data to validate our theoretical results. Note that although missing data could potentially be a limiting factor to deconfounding samples in our problem setting, when the amount of confounded data is ample, we can often assume that we will be able to deconfound enough samples to derive correct causal relationships. On the other hand, if indeed our confounded data does not contain enough samples to deconfound and intervention is not feasible on the target treatment, then one is only left with collecting new deconfounded data (potentially in additional to the confounded data). Note that in this setting our method of selecting new samples to collect can still be applied. Finally, we conclude by pointing to several promising directions for potential future research:

1. In our current model, we assume that the treatment and outcome variables are binary, and the confounders are categorical. We plan to extend our results to more general causal problems, including the cases where the causal model is linear or semi-parametric.

2. Although for confounders like genetics, the only option to estimate the causal e ect is to reveal the value of the confounder, in practice, proxies and mediators might be available for a subset of confounders. Thus, we may extend the idea of selective revelation of information beyond confounders to incorporate other variables, such as mediators and proxies.

3. Finally, our work can be extended to an adaptive setting where we can dynamically update the sample selection decision once more information about the conditional probability is revealed.

3.7 The Generalization of Our Models

3.7.1 Multiple Confounders

In this section, we show that because we do not impose any independence assumption on the set of confounder, revealing the values of all confounders o ers maximal information on the joint distribution of the confounders. In particular, we will illustrate through the case where we have two binary confounders. The extension to multiple categorical confounders is straight forward.

In the case where we have two binary confounders Z_1 and Z_2, we can express the ATE as follows:

$$\text{ATE} = \sum_{z_1;z_2} P_{Y|T;Z_1;Z_2}(1 \mid 1; z_1; z_2) * P_{Y|T;Z_1;Z_2}(1 \mid 0; z_1; z_2) \; P_{Z_1;Z_2}(z_1; z_2):$$

With an in nite amount of confounded data, we are provided with the joint distribution $P_{Y;T}(y; t)$. Thus, it remains to estimate the conditional distributions $P_{Z_1;Z_2|Y;T}$. In our paper, we consider only the non-adaptive policies, i.e., the number of samples to deconfound in each group $(y; t)$ is xed a priori. In the case where the costs of revealing the values of Z_1 and Z_2 are the same and we do not have any prior knowledge on the distributions of Z_1 and Z_2, the variables Z_1 and Z_2 becomes exchangeable. In the case where the sample selection policies are completely non-adaptive (which is the case that we consider in this paper), by the symmetry of the variables Z_1 and Z_2, we have that sampling from the joint distribution of Z_1 and Z_2 yields the maximum expected information on the value of the ATE. (Note that if the confounders take categorical values of di erent sizes and we allow adaptive policies, then we might be able to reduce the total cost of deconfounding to estimate the ATE to within a desired accuracy level.)

3.7.2 Pretreatment Covariates

In the case where we have known pretreatment covariates X, our model can be applied in estimating the individual treatment e ect where we make the common ignorability assumption

on the pretreatment covariates X and the confounder Z: given pretreatment covariates X and the confounder Z, the values of outcome variable Y, $= 0$ and $Y = 1$, are independent of treatment assignment. In this case, the distributions $P_{Y;T}.y; t/$ and $P_X.x/$ are known and the Individual Treatment E ect (ITE):

$$\begin{aligned} \text{ITE} &= \sum_{z;x} P_{Y_{T;Z;X}}.1; z; x/ * P_{Y_{T;Z;X}}.0; z; x/ \; P_{Z;X}.z; x/ \\ &= \sum_{z;x} P_{Y_{T;Z;X}}.1; z; x/ * P_{Y_{T;Z;X}}.0; z; x/ \; P_{Z_X}.z/P_X.x/: \end{aligned} \tag{3.15}$$

Note that in Eq. (3.15) the only distributions we need to estimate are the conditional distributions $P_{Z_{Y;T;X}}$. The values of $P_{Y_{T;Z;X}}$ and P_{Z_X} can be calculated from $P_{Z_{Y;T;X}}$ by first conditioning the confounded distributions $P_{Y;T}$ on the values of the pretreatment covariates X, i.e., we first subsample all confounded (outcome, treatment) pairs for a fixed value of X, say $X = x$, and then within each subsample, estimate the conditional distributions $P_{Z_{Y;T;X}}$ by applying our methods. To obtain ITE, we weight the estimates we obtain from all subsamples by $P_X.x/$.

3.8 Doubly Robust Estimator When Incorporating Confounded Data

When incorporating (in nite) confounded data, we know the confounded distribution, i.e., $P_{Y;T}.y; t/$'s (or the a_{yt}'s) are known. Thus, to obtain the maximum likelihood estimator for the outcome regression model, we need to add the constraint indicating that the estimated $P_{\hat{Y};T}.y; t/$'s, i.e., $P_{Y;T}.y; t/$'s, should equal to the known distributions a_{yt}'s. Thus, to integrate these constraints into the estimation processes for the conditional outcomes $P_{Y_{T;Z}}.1; z/$'s, we need to estimate the additional parameters $P_{Z_T}.z/$'s. Let z_i^t to denote the values of Z when $T = t$, where $i = 1; :::; N_t$ (where N_t is the total number of samples where $T = t$, and note that $N_t = \sum_z N_{tz}$). Now, we can write the constraint as

$$\sum_z P_{Y_{T;Z}}.1; z/P_{Z_T}.z/ = \frac{a_{1t}}{\sum_y a_{yt}};$$

and we will estimate the $P_{Z_T}.z/$ through the MLE estimator using regression $P_{Z_T}.z/ = .w_t u_t + b_t/$, where w_t and b_t are the weights and bias respectively, and u_t is a binary random variable that takes the value 1 if $T = t$ and 0 otherwise. Thus, to obtain the MLE estimator for the outcome regression model when incorporating confounded data, we need to solve the following systems of

constrained MLE problem:

$$(P1) \text{ for all } t, z:$$

$$\arg\max \sum_{i=1}^{N_{tz}} y_i^{tz} \log\left(\sigma(w_{tz} u_{tz} + b_{tz}) + .1 * y_i^{tz}\right) \log\left(1 * \sigma(w_{tz} u_{tz} + b_{tz})\right)$$

$$\text{s.t.} \sum_z \sigma(w_{tz} u_{tz} + b_{tz}) \sigma(w_t u_t + b_t) = \frac{a_{1t}}{\sum_y a_{yt}}$$

$$(P2) \text{ for all } t, z:$$

$$\arg\max \sum_{i=1}^{N_t} z_i^t \log\left(\sigma(w_t u_t + b_t) + .1 * z_i^t\right) \log\left(1 * \sigma(w_t u_t + b_t)\right)$$

$$\text{s.t.} \sum_z \sigma(w_{tz} u_{tz} + b_{tz}) \sigma(w_t u_t + b_t) = \frac{a_{1t}}{\sum_y a_{yt}}$$

Although we can verify that our plugin estimator used in the paper (Eq. (3.6)) is feasible to the above optimization system, a direct observation from the above optimization system (P1 and P2) is that all decision variables that share the same treatment value, t, share the same feasible region, but their objective functions are not correlated. Thus, the above optimization system is not well-defined. We conclude that extending the doubly-robust estimator to incorporate confounded data is not straight-forward and requires further research.

Checking the Feasibility of Our Estimator Recall that our plugin estimator yields the following estimated $P(Y = y, Z = z \mid t)$'s and $P(Z = z \mid t)$'s:

$$\overline{\sigma(w_{tz} u_{tz} + b_{tz})} = \frac{a_{1t} q_{1t}^z}{\sum_y a_{yt} q_{yt}^z}$$

$$\overline{\sigma(w_t u_t + b_t)} = \frac{\sum_y a_{yt} q_{yt}^z}{\sum_y a_{yt}}$$

Indeed, if we plugin $\overline{\sigma(w_{tz} u_{tz} + b_{tz})}$ and $\overline{\sigma(w_t u_t + b_t)}$ in our constraints, we have

$$\sum_z \overline{\sigma(w_{tz} u_{tz} + b_{tz}) \sigma(w_t u_t + b_t)} = \sum_z \frac{a_{1t} q_{1t}^z}{\sum_y a_{yt} q_{yt}^z} \frac{\sum_y a_{yt} q_{yt}^z}{\sum_y a_{yt}} = \sum_z \frac{a_{1t} q_{1t}^z}{\sum_y a_{yt}} = \frac{a_{1t} \sum_z q_{1t}^z}{\sum_y a_{yt}} = \frac{a_{1t}}{\sum_y a_{yt}};$$

where the last equality is because $\sum_z q_{1t}^z = \sum_z P(Z = z \mid Y = 1, T = t) = 1$.

3.9 Proofs

To begin, recall the notation introduced in § 3.3: we model the binary-valued treatment, the binary-valued outcome, and the categorical confounder as the random variables $T \in \{0, 1\}$, $Y \in \{0, 1\}$,

and $Z \in \{1, \dots, k\}$, respectively. The underlying joint distribution of these three random variables is represented as $P_{Y,T,Z}(\cdot,\cdot,\cdot)$. To save on space for terms that are used frequently, we define the following shorthand notation:

$$p^z_{yt} = P_{Y,T,Z}(y,t,z); \quad a_{yt} = P_{Y,T}(y,t); \quad q^z_{yt} = P_{Z|Y,T}(z|y,t).$$

These terms appear frequently because, to estimate the entire joint distribution on Y,T,Z (the p^z_{yt}'s), it suffices to estimate the joint distribution on Y,T (the a_{yt}'s), along with the conditional distribution of Z on Y,T (the q^z_{yt}'s): $p^z_{yt} = a_{yt}q^z_{yt}$. Finally, let $\hat{p}^z_{yt}, \hat{a}^z_{yt}$, and $\hat{q}^z_{yt}$ be the empirical estimates of p^z_{yt}, a^z_{yt}, and q^z_{yt}, respectively, using the MLE.

3.9.1 Proof of Theorem 5

Theorem 5. Using deconfounded data alone, the estimator $\widehat{ATE}(p)$ as defined in Equation (3.4) is (ϵ,δ)-close if the number of deconfounded samples is at least

$$m_{base} := C \max_{t,z} \left(\sum_y p^z_{yt}\right)^{-2} = C \max_{t,z} \frac{1}{P_{T,Z}(t,z)^2},$$

where $C := 12.5k^2 \ln(8k/\delta)\epsilon^{-2}$.

Proof. Proof of Theorem 5 This proof proceeds as follows: first, we prove a sufficient (deterministic) condition, on the errors of our estimates of p^z_{yt}'s, under which $|\widehat{ATE} - ATE|$ is small. Second, we show that the errors of our estimates of p^z_{yt}'s are indeed small with high probability.

Step 1: First, we can write the ATE in terms of the p^z_{yt}'s as follows:

$$ATE = \sum_z \left(P_{Y|T,Z}(1|1,z) - P_{Y|T,Z}(1|0,z)\right) P_Z(z) = \sum_z \left(\frac{p^z_{11}}{\sum_y p^z_{y1}} - \frac{p^z_{10}}{\sum_y p^z_{y0}}\right)\left(\sum_{y,t} p^z_{yt}\right).$$

In order for the ATE to be well-defined, we assume $\sum_y p^z_{yt} \in (0,1)$ for all t,z throughout. We can then decompose $|\widehat{ATE} - ATE|$:

$$\begin{aligned} |\widehat{ATE} - ATE| &= \left|\sum_z \left(\frac{\hat{p}^z_{11}}{\sum_y \hat{p}^z_{y1}} - \frac{\hat{p}^z_{10}}{\sum_y \hat{p}^z_{y0}}\right)\left(\sum_{y,t} \hat{p}^z_{yt}\right) - \left(\frac{p^z_{11}}{\sum_y p^z_{y1}} - \frac{p^z_{10}}{\sum_y p^z_{y0}}\right)\left(\sum_{y,t} p^z_{yt}\right)\right| \\ &\le \sum_z \left|\left(\frac{\hat{p}^z_{11}}{\sum_y \hat{p}^z_{y1}} - \frac{\hat{p}^z_{10}}{\sum_y \hat{p}^z_{y0}}\right)\left(\sum_{y,t} \hat{p}^z_{yt}\right) - \left(\frac{p^z_{11}}{\sum_y p^z_{y1}} - \frac{p^z_{10}}{\sum_y p^z_{y0}}\right)\left(\sum_{y,t} p^z_{yt}\right)\right| \end{aligned}$$

Thus, in order to upper bound $\left|\widehat{\mathrm{ATE}} * \mathrm{ATE}\right|$ by some ϵ, it suffices to show that

$$\left|\frac{\hat{p}^z_{11}}{\hat{p}^z_{y1}} * \frac{\hat{p}^z_{10}}{\hat{p}^z_{y0}} * \hat{p}^z_{yt} - \frac{p^z_{11}}{p^z_{y1}} * \frac{p^z_{10}}{p^z_{y0}} * p^z_{y;t}\right| \le \frac{\epsilon}{k}; \quad \forall z. \tag{3.16}$$

Step 2: To bound the above terms, we first derive Lemma 5 for bounding the error of the product of two estimates in terms of their two individual errors:

Lemma 5. For any $u; u' \in [*1; 1]$, and $v; v' \in [0; 1]$, suppose there exists $\epsilon \in (0; 1/$ such that all of the following conditions hold: (1) $|u' * u| \le \epsilon * (1 * |u|)$, (2) $|v' * v| \le \epsilon$, (3) $|u| + \epsilon \le 1$, (4) $|v| + \epsilon \le 1$, and (5) $\epsilon \le \min(|u|; |v|)$. Then, $|u'v' * uv| \le \epsilon$.

Proof. Proof of Lemma 5 Since $|u' * u| \le \epsilon * (1 * |u|)$, we have $u' \in [u * \epsilon * (1 * |u|); u + \epsilon * (1 * |u|)]$, and similarly, from $|v' * v| \le \epsilon$, we have $v' \in [v * \epsilon; v + \epsilon]$. Thus,

$$\begin{aligned}
|u'v' * uv| &\le \max(|u'v' * (u + \epsilon * (1 * |u|))(v + \epsilon)|, |u'v' * (u * \epsilon * (1 * |u|))(v * \epsilon)|) && \text{(because } v; v' \ge 0) \\
&= \max\left(\left|u + \epsilon * (1 * |u|)\right| * \left|v + \epsilon * (1 * |u|)\right| * \epsilon^2, \left|u + \epsilon * (1 * |u|)\right| * \left|v * \epsilon * (1 * |u|)\right| * \epsilon^2\right) \\
&= \left|u + \epsilon * (1 * |u|)\right| * \left|v + \epsilon * (1 * |u|)\right| * \epsilon^2 && \text{(because } 1 * |u| > 0) \\
&\le (|u| + \epsilon * (1 * |u|)) * |v + \epsilon * (1 * |u|)| * \epsilon && \text{(because } \epsilon^2 > \epsilon * (1 * |u|)) \\
&\le \epsilon && \text{(because } |u| + \epsilon \in [*1; 1], \text{ and } |v| + \epsilon \le 1)
\end{aligned}$$

□

We can apply Lemma 5 directly to the terms in Eq. (3.16) by setting

$$u_z = \frac{\hat{p}^z_{11}}{\hat{p}^z_{y1}} * \frac{\hat{p}^z_{10}}{\hat{p}^z_{y0}}; \quad u'_z = \frac{p^z_{11}}{p^z_{y1}} * \frac{p^z_{10}}{p^z_{y0}}; \quad v_z = \hat{p}^z_{yt}; \quad v'_z = p^z_{yt};$$

and noting that $u_z; u'_z \in [*1; 1]$, and $v_z; v'_z \in [0; 1]$. Lemma 5 implies that the upper bound in Eq. (3.16) holds if, for some $\epsilon \in (0; 1/$, we have

$$|v_z * v'_z| < \frac{\epsilon}{k} \quad \text{and} \quad |u_z * u'_z| < \frac{1 * \epsilon}{k}:$$

While we can apply standard concentration results to the $|v_z * v'_z|$ terms, the $|u_z * u'_z|$ terms will need to be further decomposed:

$$|u_z * u'_z| = \left|\frac{\hat{p}^z_{11}}{\hat{p}^z_{y1}} * \frac{\hat{p}^z_{10}}{\hat{p}^z_{y0}} * \frac{\hat{p}^z_{11}}{\hat{p}^z_{y1}} + \frac{\hat{p}^z_{10}}{\hat{p}^z_{y0}} * \frac{\hat{p}^z_{11}}{\hat{p}^z_{y1}} * \frac{\hat{p}^z_{11}}{\hat{p}^z_{y1}} + \frac{\hat{p}^z_{10}}{\hat{p}^z_{y0}} * \frac{\hat{p}^z_{10}}{\hat{p}^z_{y0}}\right|$$

It will suffice to show that for each t and z,

$$\left|\frac{\hat{p}^z_{1t}}{\hat{p}^z_{yt}} * \frac{\hat{p}^z_{1t}}{\hat{p}^z_{yt}}\right| < \frac{1 * \epsilon}{2k}: \tag{3.17}$$

Step 3: To bound these terms, we derive Lemma 6. Recall that $p_{1t}^z + p_{0t}^z, p_{1t}^z + p_{0t}^z \in (0,1)$.

Lemma 6. For any $w+s, w+s \in (0,1)$, if $(w+s, w) \succeq (w, w)$ and $(w, w) \succeq (w, w+s)$, then

$$\left|\frac{w}{w+s} - \frac{w}{w+s}\right| \leq 2.$$

Proof. Proof of Lemma 6 First, since $(w+s, w) \succeq (w, w)$, we have that

$$\frac{w+s}{w+s} \geq 1 \geq \frac{w+s}{w+s},$$

or equivalently,

$$1 - \frac{w+s}{w+s} \leq \frac{w+s}{w+s} - 1 \leq \frac{w+s}{w+s}.$$

We can apply this inequality and rearrange terms as follows to conclude the proof:

$$\begin{aligned}
&\left|\frac{w}{w+s} - \frac{w}{w+s}\right| = \left|\frac{1}{w+s}(w - w) - w\frac{w+s}{w+s}\right| \\
&\leq \frac{1}{w+s}\max\left\{0, (w - w) - \left(1 - \frac{w+s}{w+s}\right)w, (w - w) + \left(\frac{w+s}{w+s} - 1\right)w\right\} \\
&= \frac{1}{w+s}\max\left\{0, (w - w) + \frac{w+s}{w+s}w, (w - w) - \frac{w+s}{w+s}w\right\} \\
&= \max\left\{0, \frac{w - w}{w+s} + \frac{w}{w+s}, \frac{w - w}{w+s} - \frac{w}{w+s}\right\} \\
&\leq \left|\frac{w - w}{w+s}\right| + \left|\frac{w}{w+s}\right| \leq \left|\frac{w+s}{w+s}\right| + \left|\frac{w}{w+s}\right| \leq 2.
\end{aligned}$$

The second to last inequality follows from the assumption that $(w, w) \succeq (w, w+s)$. □

Lemma 6 implies that Eq. (3.17) is satisfied if

$$\left|p_{1t}^z - p_{1t}^z\right| \leq \frac{\varepsilon_y p_{yt}^z (1 - \rho)}{4k} \quad \text{and} \quad \left|p_{1t}^z + p_{0t}^z - p_{1t}^z - p_{0t}^z\right| \leq \frac{\varepsilon_y p_{yt}^z (1 - \rho)}{4k}.$$

Step 4: We've shown above that $\widehat{ATE} - ATE$ is satisfied when

$$(v_z, v_z) \leq \frac{\bar{k}}{k}, \quad \left|p_{1t}^z - p_{1t}^z\right| \leq \frac{\varepsilon_y p_{yt}^z (1 - \rho)}{4k}, \quad \text{and} \quad \left|p_{1t}^z + p_{0t}^z - p_{1t}^z - p_{0t}^z\right| \leq \frac{\varepsilon_y p_{yt}^z (1 - \rho)}{4k}, \forall t, z.$$

Note that if $\forall t, \left|p_{1t}^z + p_{0t}^z - p_{1t}^z - p_{0t}^z\right| = \left|\varepsilon_y p_{yt}^z - \varepsilon_y p_{yt}^z\right| \leq \frac{\varepsilon_y p_{yt}^z (1 - \rho)}{4k}$ then

$$(v_z, v_z) = \left(\sum_{y,t} p_{yt}^z \left(p_{yt}^z - p_{yt}^z\right), \sum_t \left(\sum_y p_{yt}^z - p_{yt}^z\right)\right) \leq \frac{\varepsilon_{y,t} p_{yt}^z (1 - \rho)}{4k} \leq \frac{(1 - \rho)}{4k}.$$

Thus, to remove the first constraint $(v_z, v_z) \leq \frac{\bar{k}}{k}$, we set

$$\frac{\bar{k}}{k} = \frac{(1 - \rho)}{4k},$$

and obtain $\rho = \frac{1}{5}$.

Step 5: To summarize so far, Lemmas 5 and 6 allow us to upper bound the error of our estimated ATE in terms of upper bounds on the error of our estimates of its constituent terms:

$$P\left(\left|\widehat{\mathrm{ATE}} - \mathrm{ATE}\right| < \epsilon\right) \geq \prod_{t,z} P\left(\left|\hat{p}^z_{1t} - p^z_{1t}\right| < \frac{\epsilon \, p^z_{yt}}{5k}\right) = \prod_{t,z} P\left(\left|\hat{p}^z_{1t} + p^z_{0t} - p^z_{1t} - p^z_{0t}\right| < \frac{\epsilon \, p^z_{yt}}{5k}\right);$$

or equivalently,

$$P\left(\left|\widehat{\mathrm{ATE}} - \mathrm{ATE}\right| \geq \epsilon\right) \leq \sum_{t,z} P\left(\left|\hat{p}^z_{1t} - p^z_{1t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) = \sum_{t,z} P\left(\left|\hat{p}^z_{1t} + p^z_{0t} - p^z_{1t} - p^z_{0t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right):$$

Applying a union bound, we have

$$P\left(\left|\widehat{\mathrm{ATE}} - \mathrm{ATE}\right| \geq \epsilon\right) \leq \sum_{t,z} P_0\left(\left|\hat{p}^z_{1t} - p^z_{1t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) 1 + P_0\left(\left|\hat{p}^z_{1t} + p^z_{0t} - p^z_{1t} - p^z_{0t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) 1: \tag{3.18}$$

Step 6: Finally, we can apply Hoeffding's inequality (Theorem 2) to obtain the upper bound for the inequality above. Let X^z_{yt} be the random variable that maps the event $Y = y, T = t, Z = z$ to $\{0, 1\}$. Then, X^z_{yt} is a Bernoulli random variable with parameter p^z_{yt}. Let m denote the total number of deconfounded samples that we have. Since p^z_{yt} is estimated through the MLE, we have $\hat{p}^z_{yt} = \frac{\sum_{i=1}^{m} X^z_{yt}}{m}$. Applying Theorem 2, we obtain:

$$P\left(\left|\frac{\sum_{i=1}^{m} X^z_{yt}}{m} - p^z_{yt}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) \leq 2\exp\left(-2m\frac{\epsilon^2 \left(p^z_{yt}\right)^2}{25k^2}\right); \text{ and} \tag{3.19}$$

$$P_0\left(\left|\frac{\sum_{i=1}^{m} X^z_{1t} + X^z_{0t}}{m} - p^z_{1t} - p^z_{0t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) 1 \leq 2\exp\left(-2m\frac{\epsilon^2 \left(p^z_{yt}\right)^2}{25k^2}\right): \tag{3.20}$$

Combining Eq.s (3.18), (3.19), and (3.20), we have

$$\begin{aligned} P\left(\left|\widehat{\mathrm{ATE}} - \mathrm{ATE}\right| \geq \epsilon\right) &\leq \sum_{t,z} P_0\left(\left|\hat{p}^z_{1t} - p^z_{1t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) 1 + P_0\left(\left|\hat{p}^z_{1t} + p^z_{0t} - p^z_{1t} - p^z_{0t}\right| \geq \frac{\epsilon \, p^z_{yt}}{5k}\right) 1 \\ &\leq 4k \max_{t,z}\left(2\exp\left(-2m\frac{\epsilon^2 \left(p^z_{yt}\right)^2}{25k^2}\right)\right) = 8k \max_{t,z}\exp\left(-2m\frac{\epsilon^2 \left(p^z_{yt}\right)^2}{25k^2}\right); \end{aligned}$$

where the second line follows from the fact that, since t is binary, there are $4k$ terms in total. Solving the above equation, we conclude that $P\left(\left|\widehat{\mathrm{ATE}} - \mathrm{ATE}\right| \geq \epsilon\right) \leq \delta$ is satisfied when the sample size m is at least

$$m \geq \frac{125k^2 \ln\left(\frac{8k}{\delta}\right)}{\epsilon^2} \max_{t,z} \frac{1}{\left(p^z_{yt}\right)^2}:$$

□

3.9.2 Proof of Proposition 7

Proposition 7. (Lower Bound with respect to $P_{Y,T,Z}$) Fix any confounded distribution and assume that in nite confounded data is given (or equivalently $P_{Y,T}$ is known). There exists $\epsilon > 0$ such that no (ϵ, δ)-close estimator exists. Speci cally, for any number of deconfounded samples[10] there exist two underlying distributions $P^1_{Y,T,Z}$ and $P^2_{Y,T,Z}$ with the following properties:

- Both of their confounded distributions are $P_{Y,T}$.
- No algorithm can correctly identify both of them with probability more than $1 - \delta$ using at most m deconfounded samples.
- Their corresponding ATE's are ϵ apart: $|\text{ATE}(P^1_{Y,T,Z}) - \text{ATE}(P^2_{Y,T,Z})| \geq \epsilon$.

Proof. Proof of Proposition 7 It su ces to show for the case where confounder takes binary value. The extension to categorical confounder is straightforward as illustrated in the proof of Theorem 9 in Ÿ 3.9.7. Let $q_{yt} = P(Z = 1 \mid Y = y, T = t)$. To show that Proposition 7 is true, it is su cient to show that there exist a positive constant c (that depends on a) such that for all fixed a, there exists a pair of q and q'' such that $|\text{ATE}(a; q) - \text{ATE}(a; q'')| > c$, with q and q'' close in distribution. We proceed by construction. For fixed a, consider the following q pairs: $q = (q_{00}, 0, q_{10}, \epsilon)$ and $q'' = (q_{00}, \epsilon, q_{10}, 0)$. Then, we have

$$\text{ATE}(a; q) = (a_{00}q_{00} + a_{10}q_{10} + a_{11}) + \frac{a_{11}(1-\epsilon)}{a_{11}(1-\epsilon) + a_{01}}(1 - a_{00}q_{00} - a_{10}q_{10} - a_{11}) - \frac{a_{10}q_{10}}{a_{10}q_{10} + a_{00}q_{00}}(a_{00}q_{00} + a_{10}q_{10} + a_{11}) + \frac{a_{10}(1-q_{10})}{a_{10}(1-q_{10}) + a_{00}(1-q_{00})}(1 - a_{00}q_{00} - a_{10}q_{10} - a_{11}),$$

and similarly, we have

$$\text{ATE}(a; q'') = \frac{a_{11}}{a_{11} + a_{01}(1-\epsilon)}(1 - a_{00}q_{00} - a_{01} - a_{10}q_{10}) - \frac{a_{10}q_{10}}{a_{10}q_{10} + a_{00}q_{00}}(a_{00}q_{00} + a_{01} + a_{10}q_{10}) + \frac{a_{10}(1-q_{10})}{a_{10}(1-q_{10}) + a_{00}(1-q_{00})}(1 - a_{00}q_{00} - a_{01} - a_{10}q_{10}).$$

In particular,

$$\lim_{\epsilon \to 0} |\text{ATE}(a; q) - \text{ATE}(a; q'')| = |a_{00}q_{00} + a_{10}q_{10} - a_{00} + a_{10}|, \tag{3.21}$$

where we can choose q_{00} and q_{10} to be 1.

On the other hand, we can show that the number of samples needed to distinguish q from q'' is at least $(1/\epsilon)$: since q and q'' are the same in two of the entries and symmetric on the rest two, to distinguish q and q'' is to distinguish a Bernoulli random variable with parameter 0 (denoting

[10] See footnote 6

this variable B_0) from a Bernoulli random variable with parameter (denoting this random variable B). Let f be any estimator of the Bernoulli random variable, and $x_1, \ldots, x_m$ be the sequence of m observations. Then we have $\max(E_{X \sim B_0^m}[f] - E_{X \sim B^m}[f]) \geq \|B_0^m - B^m\|_1 \geq \sqrt{2 \ln 2 / KL(B_0^m \| B^m)} \geq \sqrt{2 \ln 2 / m}$, where the last inequality is because when given m samples, $KL(B_0^m \| B^m) \leq 2 \ln 2 + (1 - 2\varepsilon) \ln \frac{1+2\varepsilon}{1-2\varepsilon} / m \leq 2 m$. On the other hand, any hypothesis test that takes n samples and distinguishes between $H_0 : X_1, \ldots, X_n \sim P_0$ and $H_1 : X_1, \ldots, X_n \sim P_1$ has probability of error lower bounded by $\max(P_0(1), P_1(0)) \geq \frac{1}{4} e^{-n KL(P_0 \| P_1)}$, where $P_0(1)$ indicates the probability that we identify class H_0 while the true class is H_1. Since $P_0(1) + P_1(0) \geq$, by contradiction, we can show that $m \geq \ln(\quad) / \quad$.

Note that this lower bound on m can be arbitrarily large by choosing ε to be sufficiently small. However their ATE values stay constant away as observed in Eq. (3.21). Thus, for every fixed confounded distribution encoded by a and fixed number of deconfounded samples m, we can always construct a pair of conditional distributions encoded by q and q'' such that their corresponding ATEs are constant away while the probability that we correctly identify the true conditional distribution from q and q'' is less than $1 - \quad$. In particular, $\quad = c = a_{00} + a_{10}$ in the above example. (Here, we implicitly assume that $a_{00} + a_{10}$ is strictly greater than zero, i.e., $a_{00} + a_{10} > 0$.) □

3.9.3 Proof of Theorem 6

Proposition 6. (Lower Bound with respect to ε and) Fix any confounded distribution and assume that in nite confounded data is given (or equivalently $P_{Y,T}$ is known). For any ATE estimator, there exists an underlying distribution $P_{Y,T,Z}$, whose confounded distribution $P_{Y,T}$, for which the number of deconfounded samples[11] required for the estimator to be ε /-close is at least $\quad^{-2} \log(\quad^{-1})$.

Proof. Proof of Theorem 6 Again, it suffices to show for the case where the confounder is binary. The extension to categorical confounder is straightforward as illustrated in the proof of Theorem 9 in Ÿ 3.9.7. Let $q_{yt} = P(Z = 1 \mid Y = y, T = t)$. We will proceed by construction. Consider $q = (q_{00}, q_{01}, \ldots, \quad + \quad)$ and $q'' = (q_{00}, q_{01}, \quad + \quad, \quad)$, for some small . Then

$$\mathrm{ATE}(a, q) = \frac{a_{11}(\quad + \quad)}{a_{11}(\quad + \quad) + a_{01} q_{01}} (a_{00} q_{00} + a_{01} q_{01} + a_{10} \quad + a_{11}(\quad + \quad)) + \frac{a_{11}(1 - \quad - \quad)}{a_{11}(1 - \quad - \quad) + a_{01}(1 - q_{01})} (1 - a_{00} q_{00} - a_{01} q_{01} - a_{10} \quad - a_{11}(\quad + \quad)) - \frac{a_{10}}{a_{10} + a_{00} q_{00}} (a_{00} q_{00} + a_{01} q_{01} + a_{10} \quad + a_{11}(\quad + \quad)) - \frac{a_{10}(1 - \quad)}{a_{10}(1 - \quad) + a_{00}(1 - q_{00})} (1 - a_{00} q_{00} - a_{01} q_{01} - a_{10} \quad - a_{11}(\quad + \quad)),$$

[11] This applies when the deconfounded samples are generated according to $P_{Y,T,Z}$ as in this section, or are selectively deconfounded as in the following section.

and similarly, we have

$$\begin{aligned}
ATE(a; \ddot{q}) = {} & \frac{a_{11}}{a_{11} + a_{01}q_{01}}(a_{00}q_{00} + a_{01}q_{01} + a_{10}) + a_{11} + \frac{a_{11}(1 - q_{01})}{a_{11}(1 - q_{01}) + a_{01}(1 - q_{01})}(1 - a_{00}q_{00} - \\
& a_{01}q_{01} - a_{10}) - a_{11} \cdot \frac{a_{10}}{a_{10} + a_{00}q_{00}}(a_{00}q_{00} + a_{01}q_{01} + a_{10}) + a_{11} - \\
& \frac{a_{10}(1 - q_{00})}{a_{10}(1 - q_{00}) + a_{00}(1 - q_{00})}(1 - a_{00}q_{00} - a_{01}q_{01} - a_{10}) - a_{11}.
\end{aligned}$$

Ignoring the ϵ in the denominator, we have that

$$\begin{aligned}
ATE(a; q) * ATE(a; \ddot{q}) = {} & \left(\frac{a_{11}}{a_{11} + a_{01}q_{01}} + \frac{a_{10}}{a_{10} + a_{00}q_{00}}\right)(a_{00}q_{00} + a_{01}q_{01} + a_{10}) + a_{11} \\
& * \left(\frac{a_{11}}{a_{11}(1 - q_{01}) + a_{01}(1 - q_{01})} + \frac{a_{10}}{a_{10}(1 - q_{00}) + a_{00}(1 - q_{00})}\right)(1 - a_{00}q_{00} - a_{01}q_{01} - a_{10}) - a_{11} \\
& + \frac{a_{11}^2 * a_{11}a_{10}}{a_{11} + a_{01}q_{01}} * \frac{a_{11}^2 * a_{11}a_{10}}{a_{11}(1 - q_{01}) + a_{01}(1 - q_{01})}(1 - \epsilon) \\
& + \frac{a_{10}^2 * a_{11}a_{10}}{a_{10} + a_{00}q_{00}} * \frac{a_{10}^2 * a_{11}a_{10}}{a_{10}(1 - q_{00}) + a_{00}(1 - q_{00})}(1 - \epsilon) \\
& + \left(\frac{a_{11}^2}{a_{11} + a_{01}q_{01}}\right)^2 + \left(\frac{a_{11}^2}{a_{11}(1 - q_{01}) + a_{01}(1 - q_{01})}\right)^2 + \left(\frac{a_{10}^2}{a_{10} + a_{00}q_{00}}\right)^2 + \left(\frac{a_{10}^2}{a_{10}(1 - q_{00}) + a_{00}(1 - q_{00})}\right)^2
\end{aligned} \tag{3.22}$$

Similar to the proof above, let B_1 denote the Bernoulli random variable with parameter ϵ, and let B_2 denote the Bernoulli random variable with parameter $1 - \epsilon$. Then, given m deconfounded samples, we have $KL(B_1^m \| B_2^m) = m\left(\ln\frac{\epsilon}{1-\epsilon}\right)\epsilon + m(1-\epsilon)\ln\frac{1-\epsilon}{\epsilon} = m\ln\frac{(1-\epsilon)^2}{\epsilon^2}$. Thus, we have $m \geq \frac{\ln(1/\delta)}{2}$. From Eq. (3.22), we observe that $\epsilon = 1 - ATE(a; q) * ATE(a; \ddot{q}) \geq 1 - \delta$. Combining above, we have $m \geq \frac{\ln(1/\delta)}{2}$.

□

3.9.4 Proof of Proposition 8

Proposition 8. For every confounded distribution, there exists two sets of conditional distributions Q_1 and Q_2 such that any optimal sample selection policy under (a, Q_1) is not optimal under (a, Q_2).

Proof. Proof of Proposition 8 To show Proposition 8 holds, it suffices to construct two conditional distributions sets Q_1 and Q_2 such that the corresponding optimal sample selection policies differ under particular choices of a. Similar to previous proofs, it suffices to show the case where the confounder is binary. Consider $Q_1 = \{(q_{00}, \epsilon_2, \epsilon_3, \epsilon_4), (q_{00}, \epsilon_2, \epsilon_3, \epsilon_4)\}$ and $Q_2 = \{(\epsilon_1, \epsilon_2, \epsilon_3, q_{11}), (\epsilon_1, \epsilon_2, \epsilon_3, q_{11})\}$, where the values ϵ_i and ϵ_i, $i = 1, \ldots, 4$ are known, and $\epsilon_i \neq \epsilon_i$ for some i. Moreover, q_{00} and q_{11} represent two unknown parameters to be estimated. In particular,

we will chose the values of $_i$ and $_i$ such that $\hat{ATE}(a;q)\grave{}_{q\in Q_1}$ are nonconstant functions of q_{00} and $\hat{ATE}(a;q)\grave{}_{q\in Q_2}$ are nonconstant functions of q_{11}.

Observe that the optimal sample selection policy under Q_1 is $x_1 = (1;0;0;0)$ while the optimal selection policy under Q_2 is $x_2 = (0;0;0;1)$. Thus, we complete the proof.

3.9.5 Proof of Theorems 6 and 7

Theorem 6. Incorporating (in nite) confounded data, the estimator $\hat{ATE}(a;q)$ is $(\ ;\ )$-close if the number of deconfounded samples is at least

$$m_{nsp} := C\max_{t;z}\frac{\left(\sum_y a_{yt}\right)^2}{\sum_y a_{yt}q^z_{yt}} = C\max_{t;z}\frac{P_T(t)}{P_{T;Z}(t;z)^2}; \tag{3.13}$$

where $C := 12{:}5k^2\ln(8k/\ )^{*2}$.

Theorem 7. Incorporating (in nite) confounded data, the estimator $\hat{ATE}(a;q)$ is $(\ ;\ )$-close if the number of deconfounded samples, selected under the natural selection policy (NSP) is at least:

$$m_{nsp} := C\max_{t;z}\frac{\left(\sum_y a_{yt}\right)^2}{\sum_y a_{yt}q^z_{yt}} = C\max_{t;z}\frac{P_T(t)}{P_{T;Z}(t;z)^2}:$$

Under the uniform selection policy (USP):

$$m_{usp} := C\max_{t;z}\frac{4\sum_y a^2_{yt}}{\left(\sum_y a_{yt}q^z_{yt}\right)^2} = C\max_{t;z}\frac{4\sum_y P_{Y;T}(y;t)^2}{P_{T;Z}(t;z)^2}:$$

Under the outcome-weighted selection policy (OWSP):

$$m_{owsp} := C\max_{t;z}\frac{2\left(\sum_y a_{yt}\right)^2}{\left(\sum_y a_{yt}q^z_{yt}\right)^2} = C\max_{t;z}\frac{2}{P_{Z\sigma}(z\sigma)^2}:$$

Here, $C := 12{:}5k^2\ln(8k/\ )^{*2}$.

Proof. Proof of Theorem 7 In these theorems, we derive the concentration of the $\hat{ATE}$ assuming in nite confounded data, and parametrize p^z_{yt} by $p^z_{yt} = a_{yt}q^z_{yt}$. Since under in nite confounded data, a_{yt}'s are known, and thus we only need to estimate the q^z_{yt}'s. The key di erence between Theorem 7 and Theorem 5 is that now we de ne the random variables X^z_{yt} to map the event $(Z = z\sigma Y = y;T = t)$ to $\{0;1\}\grave{}$. Thus, X^z_{yt} is distributed according to Bernoulli(q^z_{yt}). Thus, to decompose $\left(a_{1t}q^z_{1t} + a_{0t}q^z_{0t}\right)^* \left(a_{1t}q^z_{1t} + a_{0t}q^z_{0t}\right)$ we rst show the following lemma:

Lemma 7. Let $X_1;\dots;X_{x_1m}$ and $Y_1;\dots;Y_{x_2m}$ be independent random variables in [0,1]. Then for any $t > 0$, we have

$$P\left(\left|\frac{\sum_{i=1}^{x_1m}X_i - E[X_i]}{x_1m} + \frac{\sum_{j=1}^{x_2m}Y_j - E\ Y_j}{x_2m}\right| \ t + k\right) \ 2\exp\left(-\frac{2m.\ t\ +\ k/^2}{\frac{2}{x_1}+\frac{2}{x_2}}\right):$$

Proof. Proof of Lemma 7 First observe that

$$P\left(\frac{\sum_{i=1}^{x_1 m} X_i * E[X_i]}{x_1 m} + \frac{\sum_{j=1}^{x_2 m} Y_j * E\ Y_j}{x_2 m} \ge t + k\right)$$
$$= P\left(\frac{1}{x_1}\sum_{i=1}^{x_1 m} .X_i * E[X_i] + \frac{1}{x_2}\sum_{j=1}^{x_2 m} .Y_j * E\ Y_j \ge x\ m\ t + m\ k\right):$$

Now, let $Z_i = \frac{1}{x_1} X_i$ if $i \in [1; x_1 m]$, and $Z_i = \frac{1}{x_2} Y_i$ if $i \in [x_1 m+1; .x_1+x_2/m]$. Then applying Theorem 2, we have

$$P\left(\sum_{i=1}^{x_1+x_2/m} .Z_i * E[Z_i] \ge x\ m\ t + m\ k\right) \le 2\exp\left(\frac{2m^2.\ t + k\ /^2}{\sum_{i=1}^{x_1+x_2/m} .M_i * m_i /^2 1}\right) = 2\exp\left(\frac{2m.\ t + k\ /^2}{\frac{2}{x_1} + \frac{2}{x_2}}\right):$$

.

□

As de ned in Ÿ 3.3, let x_{yt} denote the percentage data we sample from the group $(y; t)$. Recall that from the proof of Theorem 5, we have

$$P\left(\widehat{ATE} * ATE \ge x\right) \le \sum_{t;z} P\left(\hat{a}_{1t} q_{1t}^z * a_{1t} q_{1t}^z \ge x \frac{\sum_y p_{yt}^z}{5k}\right) + P\left(\hat{a}_{1t} q_{1t}^z + a_{0t} q_{0t}^z * a_{1t} q_{1t}^z * a_{0t} q_{0t}^z \ge x \frac{\sum_y p_{yt}^z}{5k}\right)$$
$$= \sum_{t;z} P\left(\hat{a}_{1t} q_{1t}^z * a_{1t} q_{1t}^z \ge x \frac{\sum_y a_{yt} q_{yt}^z}{5k}\right) + P\left(\hat{a}_{1t} q_{1t}^z + a_{0t} q_{0t}^z * a_{1t} q_{1t}^z * a_{0t} q_{0t}^z \ge x \frac{\sum_y a_{yt} q_{yt}^z}{5k}\right)$$
$$= \sum_{t;z} P\left(\hat{q}_{1t}^z * q_{1t}^z \ge x \frac{\sum_y a_{yt} q_{yt}^z}{5k a_{1t}}\right) + P\left(\hat{a}_{1t} q_{1t}^z + a_{0t} q_{0t}^z * a_{1t} q_{1t}^z * a_{0t} q_{0t}^z \ge x \frac{\sum_y a_{yt} q_{yt}^z}{5k}\right)$$
$$\le 4k \max_{t;z} \left\{ 2\exp\left(-2x_{1t} m \frac{\left(\sum_y a_{yt} q_{yt}^z\right)^2}{25k^2 a_{1t}^2}\right); 2\exp\left(-2m \frac{\left(\sum_y a_{yt} q_{yt}^z\right)^2}{25k^2 \sum_y \frac{a_{yt}^2}{x_{yt}}}\right) \right\};$$

where the second to last line follows from applying Lemma 7 to the second half of the line above it.

Solving the equation above, we have

$$m \ge \frac{12.5k^2 \ln.\frac{8k}{\ }/}{2} \max_{t;z} \left\{ \frac{a_{1t}^2 \_ x_{1t}}{\left(\sum_y a_{yt} q_{yt}^z\right)^2}; \frac{\sum_y a_{yt}^2 \_ x_{yt}}{\left(\sum_y a_{yt} q_{yt}^z\right)^2} \right\} = \frac{12.5k^2 \ln.\frac{8k}{\ }/}{2} \max_{t;z} \frac{\sum_y a_{yt}^2 \_ x_{yt}}{\left(\sum_y a_{yt} q_{yt}^z\right)^2}: \quad (3.23)$$

The last equality is because $a_2^2 \_ x_2; a_1^2 \_ x_1 > 0$. Under NSP $x_{yt} = a_{yt}$. Thus, we have

$$m_{nsp} := \frac{12.5k^2 \ln.\frac{8k}{\ }/}{2} \max_{t;z} \frac{\sum_y a_{yt}}{\left(\sum_y a_{yt} q_{yt}^z\right)^2}:$$

Similarly, under USP $x_{yt} = \frac{1}{4}$, and we have

$$m_{usp} := \frac{12.5k^2 \ln.\frac{8k}{}/}{2} \max_{t;z} \frac{^3{}_y 4a_{yt}^2}{^3{}_y a_{yt} q_{yt}^z{}^2}:$$

Lastly, under OWSP $x_{yt} = \frac{a_{yt}}{2^3{}_y a_{yt}}$, and we have

$$m_{owsp} := \frac{12.5k^2 \ln.\frac{8k}{}/}{2} \max_{t;z} \frac{2.^3{}_y a_{yt}/^2}{^3{}_y a_{yt} q_{yt}^z{}^2}:$$

□

3.9.6 Proof of Theorem 8

Theorem 8. Fix any Ë .0; 1_2/ For any Ë .0; 0:5 * 2 .1 * /], there exist confounded distributions a, and Q ï Q , such that $_{owsp}.a; Q/ \_ _{nsp}.a; Q/$ is arbitrarily close to zero. In addition, for all and Q, $_{owsp}.a; Q/ w 2 _{nsp}.a; Q/$:

Proof. Proof of Theorem 8 We proceed by construction. For simplicity, we illustrate the correctness of Theorem 8 for binary confounders. The extension to the multi-valued confounder is straightforward and will be demonstrated in the proof of Theorem 9.

Consider the following example $a_{01} = a_{10} = a_{11} =$, $a_{00} = 1 * 3$, and consider the following pair of q's: q = . ; ; ; c / and q¨ = . ; ; ; / , where c w $\frac{1*}{}$ is some constant. Let Q = ^q, q¨`. Thus, to obtain an ; /-close estimate over Q, it su ces to distinguish q from q¨ with high probability. To distinguish q from q¨, it su ces to distinguish c from , and thus the optimal selection policy under Q is to allocate all samples to the last (y,t) group. Then, we have ATE.a; q/ = $\frac{c}{1+c} + \frac{.1*c\ /.1*\ /}{2*c\ *} * \frac{}{1*2}$, and ATE.a; q¨/ = $\frac{1}{2} * \frac{}{1*2}$. Thus, ATE := ðATE.a; q/ * ATE. a; q¨/ð

$$\text{ATE} = \left(\frac{1}{2} * \frac{c}{c+1} * \frac{.1*c\ /.1*\ /}{2*c\ *}\right)$$

Note that when c = $\frac{1*}{}$, ATE = 0:5 * 2 .1 * / ù 0:5. Thus, for any Ë [0; 0:5 * 2 .1 * /], there exists some c such that = ATE . Then, for any , let .Q/ denote the minimum expected number of samples that we need to distinguish q from q¨ under the best estimator and under the optimal selection policy (described above). Note that since we only need to distinguish from , this Then under NSP, the minimum number of samples that we need under the best estimator equals to $_{nsp}.a; Q/ :=$ .Q/ _ , and under OWSP, the minimum number of samples that we need under the best estimator equals to $_{owsp}.a; Q/ = 4$. (Note that $x_{yt} = .\frac{1*3}{2.1*2\ /}; \frac{1}{4}; \frac{}{2.1*2\ /}; \frac{1}{4}/$ under OWSP in this example.) Thus $_{owsp}.a; Q/ \_ _{nsp}.a; Q/ = 4$. Since in this example, is at

most 1_4 owsp.a; Q/_ nsp.a; Q/ w 1 and can be arbitrarily close to 0 as E 0. (Intuitively, the rst statement is true because when t a0t ~ ³ t a1t and a00 ù a01, it is equally important to estimate q^z_{0t}'s and q^z_{1t}'s according to the ATE expression. However, under this setup, the number of samples allocated to groups .0; t/'s decreases as $a_{0;t}$'s approach to 0 under NSP, while under OWSP, half of the deconfounded samples are always dedicated to estimate q^z_{0t}'s.)

Next, we show the last sentence in Theorem 8 is true. For any x; d < 1 and for any Q, we note that when $w_{owsp} := 2\ _{nsp}.a; Q/ \max_t {}^3\ _y a_{yt}$ also achieves P.dATE * ATEðx / < under the outcome-weighted selection policy. The reason is that when using w_{owsp} number of deconfounded samples, the number of deconfounded data allocated to each group is at least as much as those under the natural selection policy. Thus, we have $_{owsp}.a; Q/ w w_{owsp} w 2\ _{nsp}.a; Q/$, where the last inequality is because $\max_t {}^3\ _y a_{yt} < 1$.

□

3.9.7 Proof of Theorem 9

Theorem 9. (Lower Bound) Fix any Ë .0; 1_2/ and any a. Then,

$$_{nsp}.a; Q/ x \frac{C_1}{2} \max_{y;t} \frac{a_{yt}.{}^3\ _{y¨} a_{y¨t},/2}{.{}^3\ _{y¨} a_{y¨t}/2};$$

$$_{usp}.a; Q/ x \frac{C_1}{2} \max_{y;t} \frac{4a^2_{yt}.{}^3\ _{y¨} a_{y¨t},/2}{.{}^3\ _{y¨} a_{y¨t}/2};$$

$$_{owsp}.a; Q/ x \frac{C_1}{2} \max_{y;t} \frac{2a_{yt}.{}^3\ _{y¨} a_{y¨t},/2}{{}^3\ _{y¨} a_{y¨t}};$$

where $t_* = 1 * t$ and $C_1 \times .k\ * 1/^2 \ln.\ ^{*1}/\ ^{*2}$.

Proof. Consider $q = .q^z_{00}; q^z_{01}; q^z_{10}; q^z_{11}/$ where $q^1_{01} =\ , q^1_{11} =\ +\ $, and $q^z_{11} = q^z_{01} *\ \_k * 1/$ for $z = 2; :::; k$, with ${}^3\ _z q^z_{01} = {}^3\ _z q^z_{11} = 1$. We assume that $q^z_{11}; q^z_{01} Ë [\ ; 1 *\ ]$ for some suitable and for all values of Z. Similarly, we consider the $q¨$ where the entries of q^z_{01} and q^z_{11} are ipped, i.e., $q¨ = .q^z_{00}; q^z_{11}; q^z_{10}; q^z_{01}/$, for some small , where the q^z_{yt}'s are de ned above. Then,

$$ATE.a; q/ = É_z \left[0 \frac{a_{11}q^z_{11}}{{}^3\ _y a_{y1}q^z_{y1}} * \frac{a_{10}q^z_{10}}{{}^3\ _y a_{y0}q^z_{y0}} 1\right] É_{y;t} a_{yt}q^z_{yt}\Big|$$

$$= \frac{a_{11}.\ +\ /}{a_{11}.\ +\ / + a_{01}}.a_{00}q^1_{00} + a_{01}\ + a_{10}q^1_{10} + a_{11}.\ +\ // * \frac{a_{10}q^1_{10}}{a_{10}q^1_{10} + a_{00}q^1_{00}}.a_{00}q^1_{00} + a_{01}\ + a_{10}q^1_{10} +$$

$$a_{11}.\ +\ // + É_{z=2}^{k} \frac{a_{11}\ q^z_{01} * \overline{k*1}}{a_{11}\ q^z_{01} * \overline{k*1}\ + a_{01}q^z_{01}}\ a_{00}q^z_{00} + a_{01}q^z_{01} + a_{10}q^z_{10} + a_{11}\ q^z_{01} * \frac{}{k * 1}\ *$$

$$É_{z=2}^{k} \frac{a_{10}q^z_{10}}{a_{10}q^z_{10} + a_{00}q^z_{00}}\ a_{00}q^z_{00} + a_{01}q^z_{01} + a_{10}q^z_{10} + a_{11}\ q^z_{01} * \frac{}{k * 1}\ ;$$

and similarly, we have

$$\mathrm{ATE}.a;q^{\cdot\cdot}/ = \frac{a_{11}}{a_{11} + a_{01}.\quad +\quad /}.a_{00}q_{00}^{1} + a_{01}.\quad +\quad / + a_{10}q_{10}^{1} + a_{11}\ / * \frac{a_{10}q_{10}^{1}}{a_{10}q_{10}^{1} + a_{00}q_{00}^{1}}.a_{00}q_{00}^{1} + a_{01}.\quad +\quad /+$$

$$a_{10}q_{10}^{1} + a_{11}\ / + \acute{E}_{z=2}^{k} \frac{a_{11}q_{01}^{z}}{a_{11}q_{01}^{z} + a_{01}\ q_{01}^{z} * \overline{k*1}}\quad a_{00}q_{00}^{z} + a_{01}\ \ q_{01}^{z} * \frac{}{k*1}\quad + a_{10}q_{10}^{z} + a_{11}q_{01}^{z}\ *$$

$$\acute{E}_{z=2}^{k} \frac{a_{10}q_{10}^{z}}{a_{10}q_{10}^{z} + a_{00}q_{00}^{z}}\quad a_{00}q_{00}^{z} + a_{01}\ \ q_{01}^{z} * \frac{}{k*1}\quad + a_{10}q_{10}^{z} + a_{11}q_{01}^{z}$$

Ignoring the in the denominator, we have that

$$\mathrm{ATE}.a;q/ * \mathrm{ATE}.\,a;q^{\cdot\cdot}/ \grave{u}\ \frac{a_{11}}{a_{11} + a_{01}}.a_{00}q_{00}^{1} + a_{01}\quad + a_{10}q_{10}^{1} + a_{11}\ /\quad + \frac{a_{10}q_{10}^{1}.a_{01} * a_{11}/}{a_{10}q_{10}^{1} + a_{00}q_{00}^{1}}$$

$$* \mathrm{H}\acute{E}_{z=2}^{k} 0\ \frac{a_{11}\_k*1}{a_{11}q_{01}^{z} + a_{01}q_{01}^{z}}.a_{00}q_{00}^{z} + a_{10}q_{10}^{z} + .a_{01} + a_{11}/q_{10}^{z}/1\ \mathrm{I}\quad * \acute{E}_{z=2}^{k} \frac{a_{10}q_{10}^{z}.a_{01} * a_{11}/}{a_{10}q_{10}^{z} + a_{00}q_{00}^{z}/} \frac{1}{k*1}$$

$$+ \frac{a_{11}^{2}}{a_{11}\quad + a_{01}}\ ^{2} + \acute{E}_{z=2}^{k} \frac{a_{11}^{2}}{a_{11}q_{01}^{z} + a_{01}q_{01}^{z}} \frac{}{.k*1/^{2}}^{2}$$

$$= \frac{a_{11}}{a_{11}\quad + a_{01}}.a_{00}q_{00}^{1} + a_{10}q_{10}^{1}/\quad + \frac{a_{10}q_{10}^{1}.a_{01} * a_{11}/}{a_{10}q_{10}^{1} + a_{00}q_{00}^{1}}\quad * \frac{a_{11}}{k*1}\acute{E}_{z=2}^{k} 0\ \frac{a_{00}q_{00}^{z} + a_{10}q_{10}^{z}}{a_{11}q_{01}^{z} + a_{01}q_{01}^{z}}1$$

$$* \frac{1}{k*1}\acute{E}_{z=2}^{k} \frac{a_{10}q_{10}^{z}.a_{01} * a_{11}/}{a_{10}q_{10}^{z} + a_{00}q_{00}^{z}}\quad + \frac{a_{11}^{2}}{a_{11}\quad + a_{01}}\ ^{2} + \acute{E}_{z=2}^{k} \frac{a_{11}^{2}}{a_{11}q_{01}^{z} + a_{01}q_{01}^{z}} \frac{}{.k*1/^{2}}^{2} \qquad (3.24)$$

Since the second order terms inis dominated by the rst order terms in , thus to nd the highest lower bound for sample complexity in this instance is to nd the largest coe cient in front of .

Assuming that ~ k andk < 1, then the maximum of Eq. (3.24) is achieved whq$_{00}^{z}$n= q_{10}^{z} = , $q_{00}^{1} = q_{10}^{1} = 1 * k$, and$q_{01}^{z} = .1 *\ /_k*1/$, and the coe cient in front of is

$$\frac{a_{11}}{a_{11} + a_{01}}.a_{00} + a_{10}/.\frac{}{-} * \frac{k*}{1*}/\ \grave{u}\ \frac{a_{11}}{a_{11} + a_{01}}.a_{00} + a_{10}/\ 0\ \frac{1}{-} * k_{1}\ :$$

Similar to the proof of Theorem 6, we havem í . $\frac{\ln.\ ^{*}1/}{2}$/. From Eq. (3.22), we observe that = ñATE.a;q/ * ATE.$a;q^{\cdot\cdot}$/ñí . /. Combining above, we havem í . $\frac{\ln.\ ^{*}1/}{2}$/. In the case above, $\grave{u}\ \frac{a_{11}}{a_{11}+a_{01}}.a_{00} + a_{10}/\frac{1}{-}$, thus, the number of deconfounded samples needed is approximately

$$m \times \frac{\ln.\ ^{*}1/a_{11}^{2}.a_{00} + a_{10}/^{2}}{^{2}.a_{11} + a_{01}/^{2}}\ 0\ \frac{1}{-} * k_{1}^{\quad 2}:$$

Let $C_1 \times .k\ \ * 1/^{2}\ln.\ ^{*}1/\ ^{*}2$. Thenm í $\frac{C_1}{2}\frac{a_{11}^{2}.a_{00}+a_{10}^{2}/}{.a_{11}+a_{01}/^{2}}$.

If we ip the values ofq_{01}^{z} andq_{11}^{z} with the values ofq_{00}^{z} andq_{10}^{z} in both q and$q^{\cdot\cdot}$, then we havem í $\frac{C_1}{2}\frac{a_{10}^{2}.a_{01}+a_{11}/^{2}}{.a_{10}+a_{00}/^{2}}$. Note that this is because that the estimation error on ATE and ATE

is symmetric. In addition, under natural selection policy, we need at least $\frac{m}{a_{11}}$ samples; uniform selection policy, we need at least $4m$ deconfounded samples; under outcome-weighted selection policy, we need at least $2\frac{a_{11}+a_{01}}{a_{11}}m$ deconfounded samples. Combining all of the above, we obtained Theorem 9.

□

3.9.8 Proof of Theorem 10

Theorem 10. Given n confounded and m deconfounded samples, with $n = m$, $\widehat{ATE}_a(q)$ is (ϵ, δ)-close when

$$\min_{y,t,z} \frac{\left(a_{yt} q^z_{yt}\right)^2}{\frac{1}{x_{yt} m} + \frac{(q^z_{yt})^2}{n}} = \min_{y,t,z} \left(\frac{P_{T,Z}(t,z)^2}{\frac{1}{x_{yt} m} + \frac{(q^z_{yt})^2}{n}}\right) \geq 4C. \tag{3.14}$$

Here, $C := 12.5k^2 \ln(8k/\delta)/\epsilon^2$.

Proof. Proof of Theorem 10 In this theorem, we derive the concentration for $\widehat{ATE}$ under finite confounded data. The difference between Theorem 7 and Theorem 10 is that now we need to estimate a_{yt} in addition to q^z_{yt}. Thus, to decompose $(\hat{a}_{yt}\hat{q}^z_{yt} - a_{yt}q^z_{yt})$, we first derive Lemma 8.

Lemma 8

Lemma 8. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be two sequences of Bernoulli random variables independently drawn from distribution p_1 and p_2, respectively. Let $S_X = \sum_{i=1}^{n} X_i$; $S_Y = \sum_{i=1}^{m} Y_i$. Then,

$$P\left(\left|\frac{S_X}{n} - \frac{S_Y}{m} - E[S_X] - E[S_Y]\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2}{\frac{1}{m} + \frac{p_2^2}{n}}\right).$$

Proof. Proof of Lemma 8 The proof follows the proof of Hoe ding's inequality:

$$P\ S_X S_Y * E[S_X]E[S_Y] \text{ x } nmt\ = P\ \exp.aS_X S_Y * aE[S_X]E[S_Y]// \text{ x } \exp.anmt/ \tag{3.25}$$

$$\text{w} \exp.*anmt/E\,[\exp.aS_X S_Y * aE[S_X]E[S_Y]//]\ ; \qquad \text{(because of Markov's inequality)} \tag{3.26}$$

$$= \exp.*anmt/E\,[\exp.aS_X.S_Y * E[S_Y]/ + aE[S_Y].S_X * E[S_X]/]$$

$$\text{w} \exp.*anmt/E\,[\exp.a\max.S_X/.S_Y * E[S_Y]/ + aE[S_Y].S_X * E[S_X]//] \qquad (\text{because } S_X \text{ x } 0) \tag{3.27}$$

$$= \exp.*anmt/E\,[\exp.an.S_Y * E[S_Y]/ + aE[S_Y].S_X * E[S_X]//]$$

$$= \exp.*anmt/E\,[\exp.an.S_Y * E[S_Y]//]\,E\,[\exp.aE[S_Y].S_X * E[S_X]//] \qquad (\text{because } X \circ Y) \tag{3.28}$$

$$= \exp.*anmt/\sum_{i=1}^{m}\sum_{j=1}^{n} E\,[\exp.an.Y_i * E[Y_i]//]\,E\ \exp.aE[S_Y].X_j * E[X_j]//$$

$$\text{w} \exp.*anmt/\sum_{i=1}^{m} \exp_0 \frac{a^2}{8}n^2 1 \sum_{j=1}^{n} \exp_0 \frac{a^2}{8}E[S_Y]^2 1 \tag{3.29}$$

$$= \exp_0 * anmt + \frac{a^2}{8}mn^2 + \frac{a^2}{8}nm^2p_2^2 1 \qquad (\text{because the minimum is achieved at } t = \frac{4t}{n + mp_2^2}) \tag{3.30}$$

$$\text{w} \exp_0 * \frac{2mnt^2}{n + mp_2^2 1} = \exp_H * \frac{2t^2}{\frac{1}{m} + \frac{p_2^2}{n} I}\ :$$

Line (3.29) is because $Y_i * E[Y_i]$ Ë ^*$E[Y_i]; 1 * E[Y_i]/$, and thus $n.Y_i * E.Y_i//$ Ë $[*nE[Y_i]; n.1 * E[Y_i]/]$. Furthermore, $E[S_Y].X_i * E[X_i]/$ Ë $.*E[X]E[S_Y]; .1 * E[X]/E[S_Y]/$. Finally, applying Hoe ding's Lemma (Lemma 1), we obtain line (3.29). □

Now we are ready to prove Theorem 10.

Proof of Theorem 10

In this theorem, we assume that the number of confounded data is nite. Thus, instead of a_{yt}, q_{yt}^z, we have estimates of them, named $\hat{a}_{yt}$, $\hat{q}_{yt}^z$. Let n_{yt} denote the number of samples in the confounded data such that $Y = y; T = t/$. Let m_{yt}^z be the number of samples in the deconfounded data such that $.Y = y; T = t; Z = z/$. Furthermore, let $n = \sum_{y;t} n_{yt}; m = \sum_{y;t;z} m_{yt}^z$. Then, under our setup, we estimate a_{yt} and q_{yt}^z as follows: $\hat{a}_{yt} = \frac{n_{yt}}{n}$; and $\hat{q}_{yt}^z = \frac{m_{yt}^z}{\sum_z m_{yt}^z}$: Thus, following the proof of Theorem 5,

we have

$$P\left(\widehat{ATE} * ATE\right) < x \;=\; P\left(\bigcap_{t;z}\left(\sum_{i1t} p_{1t}^{z} * p_{i1t}^{z}\right) < \frac{\sum_{y} p_{yt}^{z}}{5k}\right) - \left(\bigcap_{t;z}\left(\sum_{i1t} p_{1t}^{z} + p_{0t}^{z} * p_{i1t}^{z} * p_{i0t}^{z}\right) < \frac{\sum_{y} p_{yt}^{z}}{5k}\right)$$

$$= P\left(\bigcap_{t;z}\left(\sum_{i1t} a_{1t} q_{i1t}^{z} * a_{i1t} q_{i1t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{5k}\right) - \left(\bigcap_{t;z}\left(\sum_{i1t} a_{1t} q_{i1t}^{z} + a_{0t} q_{i0t}^{z} * a_{i1t} q_{i1t}^{z} * a_{i0t} q_{i0t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{5k}\right).$$

Notice that $\left(\sum_{i1t} a_{1t} q_{i1t}^{z} + a_{0t} q_{i0t}^{z} * a_{i1t} q_{i1t}^{z} * a_{i0t} q_{i0t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{5k}$ is satis ed when both

$$\left(\sum_{i1t} a_{1t} q_{i1t}^{z} * a_{i1t} q_{i1t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{10k}; \text{ and } \left(\sum_{i0t} a_{0t} q_{i0t}^{z} * a_{i0t} q_{i0t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{10k}.$$

We have:

$$P\left(\widehat{ATE} * ATE\right) < x \;=\; P\left(\bigcap_{t;z}\left(\sum_{i1t} a_{1t} q_{i1t}^{z} * a_{i1t} q_{i1t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{10k}\right) - \left(\bigcap_{t;z}\left(\sum_{i0t} a_{0t} q_{i0t}^{z} * a_{i0t} q_{i0t}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{10k}\right)$$

$$= P\left(\bigcap_{y;t;z}\left(a_{yt} q_{iyt}^{z} * a_{yt} q_{iyt}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{10k}\right).$$

Lemma 8 suggests that

$$P\left(a_{yt} q_{iyt}^{z} * a_{yt} q_{iyt}^{z} > x\right) \le 2\exp\left(-\frac{2t^{2}}{\frac{1}{x_{yt} m} + \frac{q_{yt}^{z}/2}{n}}\right).$$

Thus, applying a union bound and Lemma 8, we have

$$P\left(\widehat{ATE} * ATE > x\right) \le \sum_{y;t;z} P\left(\left(a_{yt} q_{iyt}^{z} * a_{yt} q_{iyt}^{z}\right) < \frac{\sum_{y} a_{yt} q_{yt}^{z}}{10k}\right) \le 8k \max_{y;t;z} \exp\left(-\frac{2\left(\sum_{y} a_{yt} q_{yt}^{z}\right)^{2} x^{2}}{\left(\frac{1}{x_{yt} m} + \frac{q_{yt}^{z}/2}{n}\right)/100k^{2}}\right).$$

Simplifying the equations above, we have

$$\min_{y;t;z} \frac{\left(\sum_{y} a_{yt} q_{yt}^{z}\right)^{2}}{\left(\frac{1}{x_{yt} m} + \frac{q_{yt}^{z}/2}{n}\right)/} x \;\; \frac{50k^{2} \ln \frac{8k}{2}}{2}.$$

□

3.10 Corresponding Stories

In this section, we will provide an example for each selection method such that this particular sampling performs the worst when compared with the other two methods. For the purpose of illustration, we consider binary confounder throughout this section. To ease notation, let q_{iyt} denote q_{iyt}^{1}.

A Scenario in Which NSP Performs the Worst A drug repositioning start-up discovered that drug T can potentially cure a disease which has no known drug cure and goes away without treatments once a while. Since drug T is commonly used to treat another disease, the majority patients who has disease do not receive any treatment. Among the ones who received drug T, the start-up discovered that the health outcomes of the majority of patients have improved. The start-up proposes to bring drug T to an observational study to verify whether drug T could treat disease while not controlling for patient's treatment adherence levels. As in most cases, patient's treatment adherence levels could in uence doctors' decision of whether to prescribe drug T and whether the treatment for disease will be successful. Translating this scenario into our notations, we have $a_{01} = \_1$, $a_{10} = \_2$, $a_{11} = \_3$, and $a_{00} = 1 * \sum_{i=1}^{3} \_i$, say $a = .0.9; 0.02; 0.01; 0.07/$. Now, imagine in the clinical trial, the patients are given a drug case containing drug T such that the drug case automatically records the frequency that the patient takes the drug. Somehow we know a priori that the patients who do not have health improvement have on average poor treatment adherence, e.g. $q_{00} = 0.9; q_{01} = 0.7$; furthermore, those who have health improvement on average have good treatment adherence, e.g. $q_{10} = 0.01; q_{11} = 0.3$. Deconfounding according to NSP, i.e. $x = .a_{00}; a_{01}; a_{10}; a_{11}/$, in this case, will select most samples from the group $(Y = 0; T = 0/$. Since the ATE depends on the estimation that relies on both $T = 0$, and $T = 1$, one would expect that NSP and OWSP will outperform NSP. The left column in Fig. 3.3 con rms this hypothesis.

A Scenario in Which USP Performs the Worst A group biostatisticians discovered that mutations on gene T is likely to cause cancer Y in patients with a particular type of heart disease. In particular, they discovered that among the those heart disease patients, 79% of patients have neither mutation on T nor cancer Y; 18% patients have both mutation on T and cancer Y. In other words, $a_{00} = 0.79; a_{11} = 0.18$. Furthermore, we have $a_{01} = 0.01; a_{10} = 0.02$. This group of biostatisticians want to run a small experiment to con rm whether gene T causes cancer Y. In particular, they are interested in knowing whether those patients also have mutations on gene Z, which is also suspected by the same group of biostatisticians to cause cancer Y. Somehow, we know a priori that $q_{00} = 0.5; q_{01} = 0.01; q_{10} = 0.05; q_{11} = 0.5$. From the calculation of the ATE, it is not di cult to observe that the error on the ATE is dominated by the estimation errors on $q_{00}; q_{11}$. Thus, we should sample more from the groups $(Y = 0; T = 0/$ and $.Y = 1; T = 1/$.

A Scenario in Which OWSP Performs the Worst A team wants to reposition drug T to cure diabetes. Drug T has been used to treat a common comorbid condition of diabetes that appears in 31% of the diabetic patient population. Among those patients who receive T, drug about 97% has improved health, that is $a_{01} = 0.01$ and $a_{11} = 0.3$. Among the patients who have never received drug T, about 70% have no health improvement, that is $a_{00} = 0.5$, and $a_{10} = 0.19$.

Let $q_{00} = 0.05$, $q_{01} = 0.5$, $q_{10} = 0.055$, and $q_{11} = 0.4$. In the ATE, it is easy to observe that $\frac{a_{11}q_{11}}{a_{11}q_{11}+a_{01}q_{01}}$ and $\frac{a_{11}.1*q_{11}/}{a_{11}.1*q_{11}/+a_{01}.1*q_{01}/}$ are both dominated by 1 regardless of the estimates of q_{11} and q_{01}. In this case, USP outperforms OWSP and NSP when the sample size is larger than 200. On the other hand, the bottom figure in the third column of Fig. 3.3 shows that, when averaged over all possible values of q, OWSP performs the best.

Figure 3.2: Performance of the four sampling policies over 13,000 distributions, $P_{Y,T,Z}$, assuming infinite confounded data. Top row (left and middle): averaged absolute error over all 13,000 distributions for varying numbers of deconfounded samples. Top row (right): averaged squared error over all 13,000 distributions. Middle row: error comparison (each point is a single distribution averaged over 100 replications) for 1,200 deconfounded samples. Bottom row: variance comparison (each point corresponds to one of the 13,000 distributions and the variance is calculated over the 100 replications) between selected sample selection policies. Bottom row: each dot corresponds to the difference between the variance of a pair of selected methods under one instance and a fixed number of deconfounded samples. Bottom left: a positive y-axis value implies that USP yields a smaller variance than NSP. Bottom middle: a positive y-axis value implies that OWSP yields a smaller variance than NSP. Bottom right: a positive y-axis value represents that OWSP yields a smaller variance than USP.

Figure 3.3: Comparison of selection policies for adversarially chosen instances. Top row left: $a = .0{:}9; 0{:}02; 0{:}01; 0{:}07/$ and $q = .0{:}9; 0{:}7; 0{:}01; 0{:}3/$, where NSP performs the worst. Top row middle: $a = .0{:}79; 0{:}01; 0{:}02; 0{:}18/$ and $q = .0{:}5; 0{:}01; 0{:}05; 0{:}5/$, where USP performs the worst. Top row right: $a = .0{:}5; 0{:}01; 0{:}19; 0{:}3/$ and $q = .0{:}05; 0{:}5; 0{:}055; 0{:}4/$, where OWSP performs the worst. Bottom row: generated with the same a's but averaged over 500 q's drawn uniformly from $[0; 1]^4$.

Figure 3.4: Performance Insights of the four sampling policies over a different set of 13,000 distributions $P_{Y,T,Z}$, assuming infinite confounded data. The y-axis of all figures is the average absolute error (AAE) difference between a pair of selected methods. Each instance is averaged over 100 replications. Top row: contains 13,000 dots, each representing an instance. The number of confounded data is fixed at 1200. The x-axis is the level of confoundedness of an instance. Middle and Bottom rows: at each level of deconfounded samples, each figure contains 130 dot, each representing one confounded distribution, averaged over 100 conditional distributions. The x-axis is the number of deconfounded samples measured in steps of 100. Middle row: the color map corresponds to the level of treatment inbalance of an instance. Bottom row: the color map corresponds to the level of (maximum) outcome inbalance within each treatment group of an instance.

Figure 3.5: Experiment on nite confounded data over 13,000 distributions, each averaged over 100 replications. The number of deconfounded samples is xed at 100. Left: averaged over the 13,000 distributions. Middle and Right: error comparison at 681 confounded samples.

Figure 3.6: Performance of the four sampling policies on the COSMIC dataset assuming in nite confounded data. 275 unique (cancer, mutation, mutation) combinations were extracted. Left: averaged over 275 instances, and each averaged over 10,000 replications. Middle and Right: error comparison at 45 deconfounded samples.