Our reference: EOR18392

AUTHOR QUERY FORM

ELSEVIER	**Journal: EOR** **Article Number: 18392**	**Please e-mail your responses and any corrections to:** **E-mail: correctionsaptara@elsevier.com**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact c.david@elsevier.com immediately prior to returning your corrections.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the '**Q**' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	AU: The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.
Q2	AU: Please provide page-number in Ref. Bothwell et al. (2018).
Q3	AU: Please provide journal title/book title in Ref. Tanaka and Yamamura (2003).
	Please check this box or indicate your approval if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Highlights

• A Partially Observable Collaborative Model optimizes treatment selection. • The model estimates a personalized chronic disease progression model. • The model has superior performance in simulated chronic depression treatment. • The model recommends personalized adaptive treatment in precision medicine.

European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Partially observable collaborative model for optimizing personalized treatment selection

Jue Gong, Shan Liu*

Department of Industrial and Systems Engineering, University of Washington, United States

ARTICLE INFO

Article history:
Received 16 April 2021
Accepted 8 March 2023
Available online xxx

Keywords:
OR in medicine
Partially observable Markov decision process
Medical decision making
Disease progression modeling

ABSTRACT

Precision medicine that enables personalized treatment decision support has become an increasingly important research topic in chronic disease care. The main challenges in designing a treatment algorithm include modeling individual disease progression dynamics and designing adaptive treatment selection strategy. This study aims to develop an adaptive treatment selection framework tailored to an individual patient's disease progression pattern and treatment response. We propose a Partially Observable Collaborative Model (POCM) to capture the individual variations in a heterogeneous population and optimize treatment outcomes in three stages. The POCM first infers the disease progression models by subgroup patterns using population data in stage one and then fine-tunes the models for individual patients with a small number of treatment trials in stage two. In stage three, we show how the treatment policies based on the Partially Observable Markov Decision Process (POMDP) can be tailored to individual patients by utilizing the disease models learned from the POCM. Using a simulated population of chronic depression patients, we show that the POCM can more accurately estimate the personal disease progression than the traditional method of solving a hidden Markov model. We also compare the POMDP treatment policies with other heuristic policies and demonstrate that the POCM-based policies give the highest net monetary benefits in majority of parameter settings. To conclude, the POCM method is a promising approach to model the chronic disease progression process and recommend a personalized treatment plan for individual patients in a heterogeneous population.

1. Introduction

Personalized treatment of chronic disease is a sequence of treatments tailored to individual patient's characteristics, disease history, and treatment response. Developing a personalized treatment plan is a difficult sequential decision-making problem due to uncertainty in observing the patient's true health state, predicting disease progression, and estimating treatment response. Treatment selection for chronic diseases requiring long-term care aims at optimizing the patient's health outcome within resource limits. The main challenges include estimating personalized disease progression dynamics and optimizing treatment selection in real-time. To address these challenges, we propose a mathematical framework for optimizing the personalized treatment of a chronic disease under partially observable health conditions and uncertain treatment outcomes.

There are multiple treatment options for many chronic diseases that can be selected over a long time horizon. Treatment options may include medications, medical devices, behavioral therapies, or no further treatment. Traditionally, clinicians make treatment decisions based on their experience and expertise during outpatient visits. These decisions broadly follow published treatment guidelines and expert consensus documents established from clinical trial data at the population level (Campbell et al., 2000; Dickstein et al., 2008). In recent years, widely implemented electronic health record (EHR) systems, expanded use of clinical decision-support tools, and digital/mobile health technologies are moving clinical care into an era of precision medicine, which is defined as "treatments targeted to the needs of individual patients on the basis of genetic, biomarker, phenotypic, or psychosocial characteristics that distinguish a given patient from other patients with similar clinical presentations" (Jameson & Longo, 2015). The confluence of data science, machine learning, artificial intelligence (AI) and precision medicine generates a vibrant research area that anticipates the next revolution in healthcare. Foreseeable benefits of personalized medicine include faster, safer, cheaper, more convenient, and higher quality of care for patients. It has been used in a variety of medical field such as cancer, cardiovascular disease, metabolic

* Corresponding author.
E-mail address: liushan@uw.edu (S. Liu).

https://doi.org/10.1016/j.ejor.2023.03.014
0377-2217/

disease, psychiatry, and pharmacogenomics, etc. (Jameson & Longo, 2015).

The first challenge is modeling the patient's unique disease progression dynamics. The maximum likelihood estimation (MLE) of progression parameters from observational data is one common method. A naïve approach is to estimate a model separately for each patient, but this approach often performs poorly due to the lack of sufficient longitudinal data collected from a single patient. Therefore, cohort data is often used to identify subgroups of similar patients in which their disease progression dynamics can be represented by subgroup models, but any heterogeneity within the subgroup is ignored. Methods for discovering subgroup structures include K-means clustering, hierarchical linear modeling, growth mixture modeling, latent class analysis, latent class growth analysis, and latent class growth mixture modeling (Twisk & Hoekstra, 2012). A more advanced approach is the collaborative model (CM) (Lin, Huang, Simon, & Liu, 2016; Lin et al., 2017; Lin, Liu, & Huang, 2018; You, Liu, Byon, Huang, & Jin, 2018). The CM framework uses a set of basis models to represent patterns as subtypes of disease progression. An individual patient's progression is a weighted combination of these basis models, where the weight is called the membership.

To build an individual disease progression model, a hidden Markov model (HMM) can account for unobservable health states that represent the true disease severity; and observations are test scores that are imperfect measurements of the health states. Model parameters include transition probabilities between health states and emission probabilities of the observations given the true health states. To utilize the subgroup structure in the population data, we propose a model that combines CM and HMM, named the Partially Observable Collaborative Model (POCM). In the POCM, the basis model is an HMM for each subtype of disease progression, and the individual progression model is a weighted combination of the basis HMMs. We develop an efficient algorithm for parameter estimation and prove that the proposed algorithm can guarantee convergence to a stationary solution similar to CM. We assume that the basis models learned from longitudinal observations of existing patients can be used to estimate future patients' individual models. Therefore, we only need to learn the membership from a small number of treatment trials to form the disease model for a new patient.

The second challenge is designing an adaptive treatment selection strategy based on the patient's past treatment response. Among research in algorithmic-based treatment planning, Markov decision process (MDP), partially-observable MDP (POMDP), reinforcement learning, and multi-armed bandits are among the most widely-used tools. We propose a POMDP model for making a sequence of adaptive treatment decisions based on the estimated individual disease dynamics. POCM or an existing HMM algorithm, i.e., the Baum-Welch algorithm (Baum, Petrie, Soules, & Weiss, 1970), can be used to estimate individual transition and emission matrices, representing the patient's response to treatment. The objective is to maximize discounted total rewards measured using Net Monetary Benefit (NMB) (defined as total health benefits $\times$ willingness-to-pay - total cost). Health benefits can be measured by quality-adjusted life years (QALYs) gained.

We are interested in the question of whether POCM is a better model than HMM alone in the sense that the parameters inferred from POCM can lead to a better treatment selection policy. The POCM-based treatment strategy is developed in three stages to provide a better understanding of the disease, its prognosis, and the most effective treatment similar to the scope laid out in precision medicine (Jameson & Longo, 2015). (1) In the learning stage, the basis models representing disease progression subgroups and patients' memberships are learned from existing dataset of patients under treatment. (2) In the fine-tuning stage, the personalized disease model of a new patient is initialized using basis models estimated from the learning stage, and the membership is updated under different treatment by running separate trials. (3) In the decision stage, the optimal treatment strategy for each patient is obtained by solving a POMDP using the disease model estimated from the fine-tuning stage (Fig. 1). We compare the outcomes of the POMDP model estimated from POCM with an POMDP model estimated from HMM and several heuristic treatment selection policies. The length of learning, fine-tuning, and decision stages can be empirically determined based on data availability, the disease application, and treatment options. The only requirement is that the number of periods in each stage should be sufficient to achieve the goal (i.e., learning, experimentation, or optimization) set for that stage.

Our modeling framework is motivated by chronic depression treatment. Depression is one of the most common mental disorders in the U.S., affecting more than 10% of the population (Pratt & Brody, 2014). Depression is difficult to diagnose because symptoms can manifest in various ways and a patient's true health state is difficult to measure (e.g., no biomarker, blood test, nor brain scan can make a conclusive depression diagnosis). Screening, diagnosis and monitoring tools currently include physical exams and instruments such as the Patient Health Questionnaire (PHQ-9), Beck Depression Inventory (BDI), Center for Epidemiologic Studies-Depression Scale (CES-D), and Hamilton Rating Scale for Depression (HRSD). We believe the importance of modeling the latent disease states as partially observable via an imperfect monitoring tool, e.g., the commonly used PHQ-9. Treatment for depression includes psychotherapy, antidepressants, or a combination of the two with supportive care (Guideline Development Panel for the Treatment of Depressive Disorder, 2019). There are no clear evidence-based guidelines on when/how to switch and augment treatment for depression. Furthermore, staying on inefficient treatments may induce more cost without any benefit for patients, and lead to treatment-resistance or addiction to medications. Therefore, there is a great need to tailor treatment to the clinical profile of each patient and classify them into subpopulations that differ in their treatment response, and select treatment to those who will benefit (Phillips, 2018). Currently, machine learning and AI are increasingly developed for mental health diagnosis and treatment interventions. Some examples include sensing and digital phenotyping, natural language processing of clinical notes and social media content, and chatbots as therapeutic agent (D'Alfonso, 2020).

We demonstrate our framework using a simulated population of chronically depressed patients. The challenge for chronic depression is that the disease's natural history is not well known and varies significantly from person to person. There is a large body of literature on discovering depression trajectories (Musliner, Munk-Olsen, Eaton, & Zandi, 2016), including some of our own work (Lin et al., 2016). One known fact is that there are three to five strong progression patterns in the population. A chronically depressed person can go through periods of response, remission, relapse, and non-response on a series of treatments (Rush et al., 2006). A landmark depression treatment trial (STAR*D) tested four steps of treatment (Rush et al., 2006), which included various antidepressants alone or in a combination of cognitive therapy (a type of psychotherapy). One recent cost-effectiveness study simulated two first-line depression treatments and several follow-up treatment options based on the STAR*D trial (Ross, Vijan, Miller, Valenstein, & Zivin, 2019). One critical treatment question is when to combine the two types of therapies. Therefore, in our case study, we simulate a two-treatment selection problem; Treatment-I uses antidepressant medication only, and Treatment-II is an intensive outpatient program with additional psychotherapy and behavioral counseling (treatment augmentation). We assume treatment alters disease progression in a probabilistic way, defined by the treat-

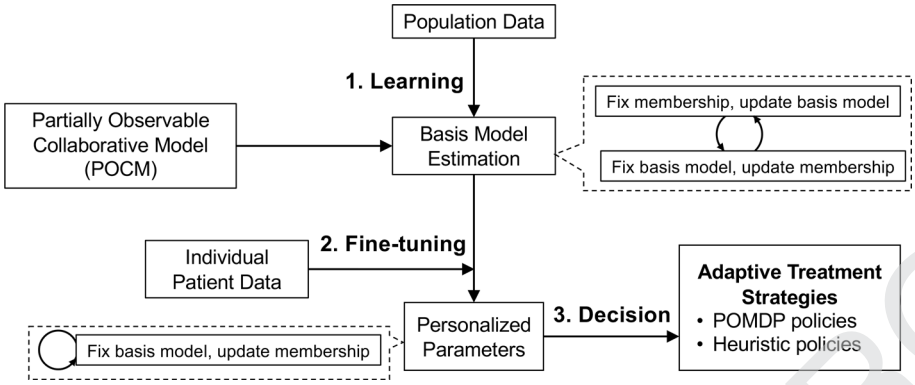

Fig. 1. Overview of the three-stage POCM framework.

ment transition matrix. We are agnostic about the types of drugs or talk therapy under Treatment-I or II. In the learning stage, the basis models are learned from a dataset of patient observations (i.e., PHQ-9 scores) on Treatment-I. The population average treatment effect of Treatment-II is assumed to be known from clinical trials or observational studies, and such knowledge is used to estimate the basis model parameters for Treatment-II.

The main contributions are twofold. First, we propose a mathematical framework, POCM, to characterize the individual variations in chronic disease progression in which the health states are partially observable. We accomplish this task through the learning stage and the fine-tuning stage. Second, we demonstrate the performance of POCM-based POMDP treatment policies compared with a set of heuristic policies in the decision stage. Through a simulated case study of chronic depression, we show that POCM can be a better model than individual HMM estimation under most parameter settings. The treatment policies' outcomes are evaluated in sensitivity analyses, including uncertainties in the treatment effect, utilities of the health states, and cost of treatments.

The remainder of the paper is structured as follows: Section 2 provides the relevant literature on optimal treatment selection models. Section 3 introduces the POCM method in the three-stage personalized treatment selection framework. The numerical results of a simulated depression case study demonstrating the performance of the proposed method are given in Section 4, followed by discussion and concluding remarks in Section 5.

2. Relevant literature

Our work is related to the literature on optimal treatment selection. MDP and POMDP have been used to determine the optimal treatment plan for many diseases, assuming that the disease transition is Markovian. Shechter, Bailey, Schaefer, & Roberts (2008) developed an MDP model to find the optimal time to initiate HIV therapy for U.S. veterans. Mason et al. (2012) used an MDP model to optimize the treatment decision for patients with type 2 diabetes. Maillart, Ivy, Ransom, & Diehl (2008) formulated a partially-observed Markov model to select effective breast cancer screening policies. Faissol, Griffin, & Swann (2007) used a POMDP to determine the best timing of treatment decisions when the disease's presence is not known in advance of hepatitis C screening. Saure, Patrick, Tyldesley, & Puterman (2012) formulated a discounted infinite-horizon MDP for scheduling cancer treatments in radiation therapy units. Otten, Timmer, & Witteveen (2020) developed a POMDP model to optimize and personalize breast cancer follow-up. Madadi, Zhang, & Henderson (2015) evaluated breast cancer mammography screening policies considering adherence behavior. Skandari & Shechter (2021) designed ongoing treatment plans for a heterogeneous population in disease progression and response to medical interventions using a POMDP model that learns the patient type. Other methods include the Kalman filter (Helm, Lavieri, Van Oyen, Stein, & Musch, 2015; Kazemian, Helm, Lavieri, Stein, & Van Oyen, 2019), multi-arm bandit (Ayer, Zhang, Bonifonte, Spaulding, & Chhatwal, 2019; Negoescu, Bimpikis, Brandeau, & Iancu, 2017), and mixed-integer programming (Chen, Ayer, & Chhatwal, 2018). A key distinction between these research and our settings is that this stream of work mainly focuses on finding the optimal treatment based on heterogeneous population characteristics, while our study focuses on online treatment selection optimized for individual patients.

Our work also relates to the stream of research that focuses on the personalization of treatment selection. Wang, Sontag, & Wang (2014) proposed a personalized disease progression model by combining Markov jump process and Markov chains. Schulam & Saria (2015) proposed a hierarchical latent variable model that individualizes disease trajectories predictions and provided the algorithm for learning population, subpopulation, and individual parameters. Ayer, Alagoz, & Stout (2012) designed a personalized mammography screening policy based on the prior screening history and personal risk characteristics of women. Lavieri, Puterman, Tyldesley, & Morris (2012) proposed an individual disease progression of prostate cancer patients using the dynamic Kalman filter model to estimate the individual parameters from the population characteristics. These papers focus on a target disease and require domain knowledge on the progression of the target disease. Our purpose is different: we propose an offline and online learning method to estimate a personalized chronic disease model based on longitudinal observations. Our method extends the CM (Lin et al., 2016; Lin et al., 2017; Lin et al., 2018) method with the ability to model partially-observable disease conditions by adding latent variables to represent health states. MLE is the most common approach of parameter inference from observational data in disease models. However, the inference of latent variables with MLE is usually difficult due to the nonconvexity of the likelihood function. The Expectation-Maximization (EM) algorithm can overcome this difficulty by iteratively estimating the intermediate states and maximizing the approximate likelihood based on the intermediate states (Dempster, Laird, & Rubin, 1977; Louis, 1982). We develop a variant of the EM algorithm for the inference of CM for diseases with latent health states.

Finally, our work relates to the stream of research that applies AI methods to medical decision making (MDM) problems. A recent review of supervised and unsupervised learning applications in MDM is Jiang et al. (2017). Reinforcement learning (RL), which formulates the process of an agent (e.g., the decision maker) interacting with an environment (e.g., the disease progression model), is widely used for sequential decision-making problems in health-

Table 1
List of Notations.

a, s, o	Action, health state, observation
$\mathcal{A}, \mathcal{S}, \Omega$	Action space, state space, observation space
$\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{\pi}_k$	Basis transition matrix, emission matrix, initial state distribution of group k
$\mathbf{A}_i^g, \mathbf{B}_i^g, \boldsymbol{\pi}_i^g$	Ground truth transition matrix, emission matrix, initial state distribution for patient i
$\hat{\mathbf{A}}_i, \hat{\mathbf{B}}_i, \hat{\boldsymbol{\pi}}_i$	Individual transition matrix, emission matrix, initial state distribution
θ	Set of basis parameters
$\mathbf{C}$	Membership
$\mathbf{O}$	Observations of all patients
$\mathbf{S}$	Latent states of all patients
$\mathbb{S}$	The set of all possible combinations of $\mathbf{S}$
K	Number of basis models
$\mathbf{W}$	Similarity matrix
μ	Tuning parameter in POCM objective
$\gamma_{i,t}(s)$	The probability that patient i is at state s at time t
$\xi_{i,t}(s, s')$	The probability that patient i is at state s at time t, and at state s' at time $t+1$
$\boldsymbol{\pi}^{(m)}, \mathbf{A}^{(m)}, \mathbf{B}^{(m)}, \mathbf{C}^{(m)}$	Estimated parameters after m iterations
$b(s)$	Belief of health state s
T	Total number of periods in each stage
$\kappa(s)$	Utility for health state s
$r(s, a)$	Reward of one period when the true health state is s and the action is a
$h(a)$	Cost of treatment a
λ	Willingness-to-pay ($/QALY)
ρ	Treatment effect factor
ϕ	Discount factor in POMDP
ζ^2	Control parameter for the significance of latent structure in the population

care. Ayer (2015) proposed an inverse RL to identify the optimal screening strategies for breast cancer in the setting of a partially observable environment. In order to simultaneously utilize the biomedical dynamics across multiple patients, Lee, Lavieri, Volk, & Xu (2015) designed three classes of RL policies for the screening of hepatocellular carcinoma. In addition, RL can be used to solve MDP and POMDP problems approximately, which is useful when the state space is large and computation resource is limited (Jaakkola, Singh, & Jordan, 1995; Zhu, Li, Poupart, & Miao, 2018). Otherwise, the exact solution via dynamic programming may be preferred. In addition, RL requires a large number of iterations of interactions (typically larger than 10,000) with the environment to ensure convergence of the optimal policy, while dynamic programming does not require interaction with the environment before performing the policies. The high cost of treatment and the long interval between treatments of the MDM problems limit the application of RL. A related field of research is multi-task RL, where an agent has multiple learning tasks during one's lifetime to maximize the total reward (Tanaka & Yamamura, 2003). The agent solves multiple MDPs or POMDPs under certain distributional assumptions (Li, Liao, & Carin, 2009). The goal is to utilize the learning experience gained among different RL tasks and their similarities to improve future learning performance for every single task (Calandriello, Lazaric, & Restelli, 2014). Our work does not make any distributional assumptions for different POMDPs in the model learning process.

3. Model formulation

We describe the disease model in Section 3.1. The POCM formulation and the algorithm for parameter estimation are provided in Section 3.2. In Section 3.3, we present the POMDP model to optimize treatment selection.

3.1. Disease model

We formulate the treatment decision process as a finite-state, finite horizon, discrete-time POMDP, where the underlying HMM represents the progression dynamics of a patient's health state. Decision makers such as patients and clinicians aim to maximize the expected total discounted NMBs. We assume that the decision maker is risk-neutral. Please refer to Table 1 for a list of notations.

Decision epoch. Treatment decisions are made at a finite and discrete set of time periods $t = 1, 2, \ldots, T, T < \infty$. We assume each period is long enough for the treatment to affect disease progression probabilities. Thus, one time period is the minimum time that a treatment switch can take place. In the case of chronic depression, the treatment decision can be made monthly (Ross et al., 2019).

State and observation. Let s_t denote the health state at time t. The state space $\mathcal{S}$ is the set of all possible health states. We omit death as a health state because the goal of the model is to learn treatment response parameters for patients who are alive in both stage 2 and 3. Furthermore, the probability of death can only be estimated using population-level data, not for an individual patient (i.e., death is a binary outcome). We assume the true health state cannot be observed. Instead, at each time period, the patient's health condition is measured by an imperfect test. The observations can be either continuous variables or categorical variables depending on the disease application. In this paper, we focus on categorical observations. Let o_t denote the observation at time period t, and the observation space Ω is a finite set of all possible observation values. Since the true health state s_t is hidden, we can only maintain an estimation of the probability distribution of the state over the state space $\Delta(\mathcal{S}) = \left\{b(s_t) \in [0, 1], s_t \in \mathcal{S} \middle| \sum_{s_t \in \mathcal{S}} b(s_t) = 1\right\}$, which is usually called the belief of states.

Action. Actions include the selection of treatment types. Let a_t denote the action taken at time t. At the beginning of each time period, an action is selected with the current policy. We demonstrate the problem with two treatment types; Treatment-I is less expensive and less effective than Treatment-II. The action space $\mathcal{A}$ is the set of all possible actions, i.e. $a_t \in \mathcal{A} = \{\text{I}, \text{II}\}$. A policy $\pi : \Delta(\mathcal{S}) \to \mathcal{A}$ is the probability of taking action a when the belief of states is $\Delta(\mathcal{S})$. At the beginning of each decision epoch, the treatment type is selected by the policy. The treatment type and the corresponding observation can provide the clinician with valuable information about the patient's true health state, which in turn helps the clinician select the next period treatment.

System Dynamic. The state transition probabilities describe the disease progression under different treatment types. It is defined

as the probability that the patient will be in state $s' \in \mathcal{S}$ at time $t+1$, given that she is in state s and action a is taken, denote as $\mathbf{A}(s,a,s') = \Pr(s_{t+1} = s' | s_t = s, a_t = a)$. The emission probability is the probability of making an observation $o \in \Omega$ at time t when the true health state is $s \in \mathcal{S}$, denote as $\mathbf{B}(s,o) = \Pr(o_t = o | s_t = s)$. We assume this probability is independent of the action. We name $\{\mathbf{A}(s,a,s')\}_{a\in\mathcal{A},s\in\mathcal{S},s'\in\mathcal{S}} \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{S}|\times|\mathcal{S}|}$ as the transition probability matrix, and $\{\mathbf{B}(s,o)\}_{s\in\mathcal{S},o\in\Omega} \in \mathbb{R}^{|\mathcal{S}|\times|\Omega|}$ as the emission probability matrix.

Reward. The reward includes both health outcomes and economic costs associated with treatment. Health outcomes are measured in QALYs. Costs may include medication and outpatient service expense. QALY is a common metric to quantify the quality-of-life gains from medical interventions. One QALY represents a patient living in perfect health for one year. The immediate reward $r(s_t, a_t)$ is the NMB of treatment in one period when the patient's true health state is s_t and the action selected is a_t,

$$r(s_t, a_t) = \lambda\kappa(s_t) - h(a_t) \tag{1}$$

where $\kappa(s_t)$ is the utility of being in state s_t measured in QALYs; $h(a_t)$ is the cost of treatment a_t; λ is the willingness-to-pay (WTP) which assigns a monetary value to a QALY; \$50,000 per QALY is commonly used for WTP in the literature (Neumann, Cohen, & Weinstein, 2014).

3.2. Partially observable collaborative model (POCM)

POCM is used in the learning and fine-tuning stages. We first discover subgroup structures in the disease progression with the CM idea (Lin et al., 2016; Lin et al., 2017; Lin et al., 2018). In short, the CM assumes that a basis model represents a subtype of disease progression pattern in the population. Each individual model is a weighted combination of the basis models. The weights (memberships) capture the interpersonal variations. We assume there are K basis models and N patients with $N \gg K$. In POCM, the underlying disease dynamic is an HMM, with the basis transition matrix $\mathbf{A}_k$, the basis emission matrix $\mathbf{B}_k$, and the basis initial distribution of state $\boldsymbol{\pi}_k$ for group $k \in \{1,\dots,K\}$. We denote $\theta = \{\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{\pi}_k, k = 1,\dots,K\}$ as the basis parameters of POCM. Each patient's individual progression model, which is also an HMM, is assumed to be a linear combination of the basis models. The weight of the linear combination is called the membership vector $\mathbf{C}_i \in \mathbb{R}^K$ for patient i; we denote $\mathbf{C} = [\mathbf{C}_1,\dots,\mathbf{C}_N] \in \mathbb{R}^{N\times K}$ as the membership matrix. The individual parameters for patient i can be described as (1) initial distribution of state $\hat{\boldsymbol{\pi}}_i = \sum_{k=1}^K c_{ik}\boldsymbol{\pi}_k$, (2) transition probability matrix $\hat{\mathbf{A}}_i = \sum_{k=1}^K c_{ik}\mathbf{A}_k$, and (3) emission probability matrix $\hat{\mathbf{B}}_i = \sum_{k=1}^K c_{ik}\mathbf{B}_k$. We denote the observations of each patient $\mathbf{O}_i = [o_{i,1}, o_{i,2}, \dots, o_{i,T}] \in \mathbb{R}^T$, and the observations of all patients as $\mathbf{O} = [\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_N] \in \mathbb{R}^{N\times T}$. We denote $s_{i,t}$ as the true health state for patient i at time period t, $\mathbf{S}_i = \{s_{i,1}, s_{i,2}, \dots, s_{i,T}\} \in \mathbb{R}^T$ as the series of true health states for patient i, and $\mathbf{S} = \{\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_N\} \in \mathbb{R}^{N\times T}$ is the set of true states for all patients. We denote $\mathbb{S}$ as the set of all possible combinations of $\mathbf{S}$. The objective of the POCM is to optimize the maximum likelihood estimator of the observed sequence $\mathbf{O}$ with

$$\max_{\theta,\mathbf{C}} \quad f(\theta,\mathbf{C}) = \log\Pr\left(\mathbf{O}\middle|\theta,\mathbf{C}\right) - \frac{\mu}{2}\sum_{i,j} w_{ij}\left\|\mathbf{C}_i - \mathbf{C}_j\right\|^2 \tag{2a}$$

$$\text{s.t.} \quad \sum_{s'} \mathbf{A}_k(s,s') = 1,\ s = 1,\dots,S,\ k = 1,\dots,K \tag{2b}$$

$$\sum_{o} \mathbf{B}_k(s,o) = 1,\ s = 1,\dots,S,\ k = 1,\dots,K \tag{2c}$$

$$\sum_{s} \boldsymbol{\pi}_k(s) = 1,\ k = 1,\dots,K \tag{2d}$$

$$\sum_{k} c_{i,k} = 1,\ i = 1,\dots,N$$

$$\mathbf{A}_k(s,s'), \mathbf{B}_k(s,o), \boldsymbol{\pi}_k(s), c_{i,k} \geq 0,\ s,s' = 1,\dots,S,\ k = 1,\dots,K,\ i = 1,\dots,N \tag{2e}$$

The first four constraints guarantee that the transition probability, emission probability, initial state distribution, and membership vector sum up to one. Note the last term of the objective function is the regularization term that incorporates the similarity between patients. The regularization coefficient μ is a tuning parameter to control the importance of similarity in the objective function. The value of μ can be calibrated to achieve the maximum log-likelihood. $\mathbf{W} \in \mathbb{R}^{N\times N}$ is the similarity matrix. Details are in Appendix E.3. The similarity between two individuals can be quantified by comparing their profiles of the covariates, such as the demographics, social-economical, genetic and imaging information (Lin et al., 2016; Lin et al., 2017; Lin et al., 2018). We can simplify the regularization term as

$$\frac{1}{2}\sum_{i,j} w_{ij}\left\|\mathbf{C}_i - \mathbf{C}_j\right\|^2 = \sum_i \left(\sum_j w_{ij}\right)\mathbf{C}_i\mathbf{C}_i^\top - \sum_{i,j} w_{ij}\mathbf{C}_j\mathbf{C}_i^\top = \mathrm{Tr}\left(\mathbf{C}^\top\mathbf{L}\mathbf{C}\right),$$

where $\mathbf{L}$ is the Laplacian matrix of w_{ij}, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, and $\mathbf{D}$ is a diagonal matrix with elements $d_{ii} = \sum_j w_{ij}$.

The EM algorithm is a standard approach for inference of latent variables. An example is the Baum-Welch algorithm for the HMM inference. In the EM algorithm, in each iteration $m = 1, 2, \dots$, we estimate the latent states and maximize the likelihood based on the latent states. Computing the likelihood of observed sequence with latent variables is computationally intractable. Instead, we can replace the likelihood with an equivalent function Q, defined as

$$Q\left(\theta,\mathbf{C}\middle|\theta^{(m)},\mathbf{C}^{(m)}\right) := \sum_{\mathbf{S}\in\mathbb{S}} \Pr\left(\mathbf{O},\mathbf{S}\middle|\theta^{(m)},\mathbf{C}^{(m)}\right)\log\left[\Pr\left(\mathbf{O},\mathbf{S}\middle|\theta,\mathbf{C}\right)\right], \tag{3}$$

where $\theta^{(m)}, \mathbf{C}^{(m)}$ is the estimation of the parameters $\theta, \mathbf{C}$ after m iterations (Bishop, 2006, §9). We can solve the POCM by maximizing $Q(\theta,\mathbf{C}|\theta^{(m)},\mathbf{C}^{(m)})$ through updating θ and $\mathbf{C}$.

Lemma 1. *The following two objective functions are equivalent*

$$\arg\max_{\theta,\mathbf{C}} \Pr\left(\mathbf{O}|\theta,\mathbf{C}\right) = \arg\max_{\theta,\mathbf{C}} \sum_{\mathbf{S}\in\mathbb{S}} \Pr\left(\mathbf{O},\mathbf{S}\middle|\theta^{(m)},\mathbf{C}^{(m)}\right)\log\left[\Pr\left(\mathbf{O},\mathbf{S}\middle|\theta,\mathbf{C}\right)\right].$$

The proof of Lemma 1 is based on the proof of the equivalent objective in the general EM algorithm in (Bishop, 2006, §9). We defer the detailed proof to Appendix A. This optimization problem can be solved with the Lagrangian multiplier method. First, the Lagrangian is

$$\begin{aligned} L\left(\theta,\mathbf{C}\middle|\theta^{(m)},\mathbf{C}^{(m)}\right) = {} & Q\left(\theta,\mathbf{C}\middle|\theta^{(m)},\mathbf{C}^{(m)}\right) - \mu\mathrm{Tr}\left(\mathbf{C}^T\mathbf{L}\mathbf{C}\right) \\ & - \sum_{k=1}^K \lambda_k^{(\boldsymbol{\pi})}\left(\sum_{s=1}^{|\mathcal{S}|}\boldsymbol{\pi}_k(s) - 1\right) \\ & - \sum_{k=1}^K\sum_{s=1}^{|\mathcal{S}|}\lambda_{s,k}^{(\mathbf{A})}\left(\sum_{s'=1}^{S}\mathbf{A}_k(s,s') - 1\right) \\ & - \sum_{k=1}^K\sum_{s=1}^{|\mathcal{S}|}\lambda_{s,k}^{(\mathbf{B})}\left(\sum_{o=1}^{|\Omega|}\mathbf{B}_k(s,o) - 1\right) \end{aligned}$$

$$-\sum_{i=1}^{N}\lambda_i^{(\mathbf{C})}\left(\sum_{k=1}^{K}c_{i,k}-1\right), \tag{4}$$

where $\lambda_k^{(\boldsymbol{\pi})}$, $\lambda_{s,k}^{(\mathbf{A})}$, $\lambda_{s,k}^{(\mathbf{B})}$ and $\lambda_i^{(\mathbf{C})}$ are dual variables. The optimization problem can be simplified as maximizing the Lagrangian L by repeating the following steps for iteration $m = 1, 2, \ldots$ until convergence.

1. Fix $\mathbf{C}^{(m)}$, set $\theta^{(m+1)} = \arg\max_\theta Q(\theta, \mathbf{C}|\theta^{(m)}, \mathbf{C}^{(m)})$;
2. Fix $\theta^{(m+1)}$, set $\mathbf{C}^{(m+1)} = \arg\max_{\mathbf{C}} Q(\theta, \mathbf{C}|\theta^{(m+1)}, \mathbf{C}^{(m)}) - \mu\mathrm{Tr}(\mathbf{C}^\top\mathbf{L}\mathbf{C})$.

The inference of the POCM is an EM algorithm, where the E-step is to compute the intermediate states $\gamma_{i,t}^{(m)}(s)$ and $\xi_{i,t}^{(m)}(s,s')$ using the forward-backward algorithm (Appendix B.3), and the M-step is to maximize $Q\left(\theta, \mathbf{C}\middle|\theta^{(m)}, \mathbf{C}^{(m)}\right)$ by updating the basis model and the memberships separately. The POCM algorithm is summarized in Algorithm 1.

Algorithm 1: The POCM parameter inference algorithm.

Data: observations on N individuals $\mathbf{O}_1, \ldots, \mathbf{O}_N$; initial values for the parameters $\mathbf{C}^{(0)}, \theta^{(0)}$; similarity matrix $\mathbf{W}$; regularization coefficient μ; number of basis models, K; stopping criteria ϵ

Result: Estimator of basis model θ^* and membership $\mathbf{C}^*$

Initialize $\mathbf{C}^{(0)}, \theta^{(0)}$
for $m \leftarrow 1, \ldots$ *until converge* **do**
  E-step: Compute intermediate states $\gamma_{i,t}^{(m)}(s)$ and $\xi_{i,t}^{(m)}(s,s')$ using the forward-backward algorithm;
  M-step: Fix $\mathbf{C}^{(m)}$, set $\theta^{(m+1)} = \arg\max_\theta Q\left(\theta, \mathbf{C}|\theta^{(m)}, \mathbf{C}^{(m)}\right)$;
  M-step: Fix $\theta^{(m+1)}$, set $\mathbf{C}^{(m+1)} = \arg\max_{\mathbf{C}} Q\left(\theta, \mathbf{C}|\theta^{(m+1)}, \mathbf{C}^{(m)}\right) - \mu\mathrm{Tr}(\mathbf{C}^\top\mathbf{L}\mathbf{C})$;
  if *all elements of* $\left|\nu^{(m+1)} - \nu^{(m)}\right|$ *is less than* ϵ, *where* $\nu \in \{\boldsymbol{\pi}, \mathbf{A}, \mathbf{B}, \mathbf{C}\}$ **then**
    break;
  end
end

Theorem 1. *The basis parameters and the membership converge to an optimal solution under the iterative updating rule in Algorithm 1.*

We present the detailed derivation of the updating rule in POCM and the proof of Theorem 1 in Appendix B and C, respectively.

3.3. Adaptive decisions

In the decision stage, the clinician selects treatment for each individual patient based on the policy in each period. In the POMDP model, the policy is derived by maximizing the total expected reward

$$R = \sum_{t=1}^{T}\phi^t r(s_t, a_t), \tag{5}$$

where ϕ is the discount factor. When new observation $o_{t+1} = o$ is obtained after taking action $a_{t+1} = a$, we can update the belief by the Bayes' rule,

$$b_{t+1}(s') = \frac{\mathbf{B}(s', o)\sum_{s\in\mathcal{S}}\mathbf{A}(s, a, s')b_t(s)}{\sum_{s'\in\mathcal{S}}\mathbf{B}(s', o)\sum_{s\in\mathcal{S}}\mathbf{A}(s, a, s')b_t(s)}, \forall s' \in \mathcal{S}. \tag{6}$$

Treatment is selected with respect to a policy $\pi : \Delta(\mathcal{S}) \to \mathcal{A}$, where $\Delta(\mathcal{S})$ denotes the continuous set of probability distributions over $\mathcal{S}$, i.e., $a_t = \pi(b_t)$. We define value function of a policy π as $V^\pi : \Delta(\mathcal{S}) \to \mathbb{R}$, which is the expected discounted total reward when following policy π starting from belief b

$$V^\pi(b) = \mathbb{E}_\pi\left[\sum_{t=0}^{T}\phi^t r(b_t, \pi(b_t))\middle|b_0 = b\right]. \tag{7}$$

where $r(b_t, \pi(b_t)) = \sum_{\mathbf{S}\in\mathcal{S}} r(s, \pi(b_t))b_t(s)$. The optimal value function $V^*(b) = V^\pi(b)$ is the best value function that can be achieved with an optimal policy π^*. The Bellman equation describes the fundamental relation between V_t and V_{t+1}:

$$V_t(b) = \max_a \sum_{s_t\in\mathcal{S}} b(s_t)\Bigg(r(s_t, a) + \phi \sum_{o_{t+1}\in\Omega}\mathbf{A}(s_t, a, s_{t+1}) \sum_{s_{t+1}\in\mathcal{S}}\mathbf{B}(s_{t+1}, o_{t+1})V_{t+1}(b_{t+1}(s_{t+1}))\Bigg). \tag{8}$$

The value functions in finite-horizon POMDP are piecewise-linear and convex with respect to the belief, and can be represented as

$$V_t(b) = \sum_{s\in\mathcal{S}} r(s, a_t)b_t(s) = b_t^\top\alpha_t^a \tag{9}$$

where $\alpha_t^a \in \mathbb{R}^{|\mathcal{S}|}$ is a set of support vectors. At period t, when the belief is b_t, the optimal action is $a_t^* = \arg\max_{a\in\mathcal{A}} b_t^\top\alpha_t^a$. We can construct the support vector set $\left\{\alpha_t^a\right\}_{t=1}^T$ backward from $t = T$ to 1. In this paper, we use incremental pruning to accelerate the support vector enumeration (Cassandra, Littman, & Zhang, 1997). We include details of the algorithm in Appendix D.

4. Simulation experiment

Due to the lack of available dataset containing both longitudinal observations of depression severity and ground truth depression states, we simulate a hypothetical patient population undergoing chronic depression treatment with latent subgroup disease progression patterns. Under the three-stage framework, offline learning is done in stage 1, and online learning is done in stage 2 and 3. Therefore, we cannot use existing patient dataset due to active treatment assignment in the online stages. We describe the ground truth data generation process in Appendix E.1, the simulation process for stages 1 and 2 in Section 4.1, and the 12 treatment policies under consideration during stage 3 in Section 4.2.

We assume three health states of chronic depression; healthy (H), mild depression (M), and severe depression (S), in ascending severity of depression, i.e., $\mathcal{S} = \{\mathrm{H}, \mathrm{M}, \mathrm{S}\}$. We also assume the disease progression process is Markovian, which is commonly used to model chronic depression in the literature and tested using patient-level electronic record data (Lin et al., 2018; Ross et al., 2019). At each time period, patients take the PHQ-9 to evaluate their depression severity (Kroenke & Spitzer, 2002). The PHQ-9 has a score ranging from 0 to 27, which can be categorized into three levels, where scores $0 \sim 4$ (P1) indicate healthy to mild depression, scores $5 \sim 9$ (P2) indicate mild to moderate, and scores $10 \sim 27$ indicate major depression, i.e., $\Omega = \{\mathrm{P1}, \mathrm{P2}, \mathrm{P3}\}$ (Fig. 2).

The utility of each health state is denoted $\kappa(\mathrm{H}) \in (0,1), \kappa(\mathrm{M}) \in (0,1), \kappa(\mathrm{S}) \in (0,1)$, and $\kappa(\mathrm{H}) > \kappa(\mathrm{M}) > \kappa(\mathrm{S})$, by the assumption that more severe depression state will lead to lower health utility. The values of these utilities can be estimated using health utility elicitation methods and may vary between studies from different regions and different populations. We assume the monitoring decision epoch is 1 month (Ross et al., 2019).

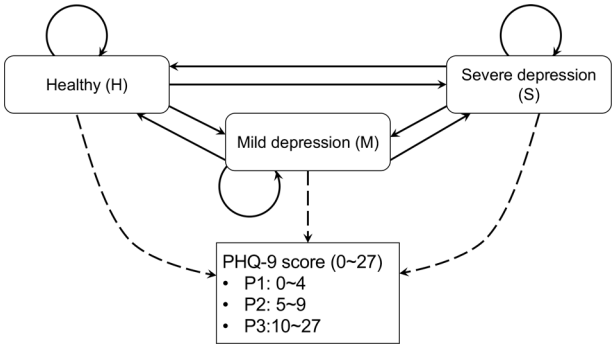

Fig. 2. Chronic depression state and observation transition diagram.

4.1. Simulation of the learning stage and the fine-tuning stage

We generate the membership vector for $N = 1000$ patients in a population with subgroup structure following the approach in Lin et al. (2018). The numerical experiment is set up to imitate real-life settings. In real-life implementation of POCM, we need a way to estimate similarity. Since patients' profiles of covariates are observable in clinical settings and often predictive of disease progression, they are good estimators for similarity. In our numerical experiment, the covariates were generated using the memberships and thus indirectly relate to disease progression patterns. The memberships are correlated with simulated patients' age and gender since these factors correlate with depression severity (Appendix E.1 and E.2). The subgroup structure is controlled by the parameter ζ^2 such that a larger magnitude of ζ^2 corresponds to a more significant subgroup structure. We randomly split the cohort into training patients ($N^{\text{train}} = 800$) and testing patients ($N^{\text{train}} = 200$).

In stage 1, the purpose of POCM training is to estimate the K basis models. The number K can be determined by expert domain knowledge or the Akaike information criterion (AIC) during model selection (Lin et al., 2016; Lin et al., 2017; Lin et al., 2018). The AIC balances the number of estimated parameters in the model and the model fit. The preferred model with the best K is the one that minimizes $\text{AIC} = 2k - 2\ln(L)$, where small k is the number of estimated parameters in the model, and L is the maximum value of the likelihood function for the model. For the HMM comparison, we estimate a unique HMM for each of the K groups. The groups are assigned using K-means clustering (Lloyd, 1982) over the patient profile of covariates (Appendix E.3). Both POCM and HMM estimations use the training patients' PHQ-9 scores which are generated for 20 periods under Treatment-I from the ground truth. Since the estimation of POCM and HMM involves the EM algorithm, we need to train the model with different initial values to avoid local optimum. We can use a generic procedure that many sets of initial values are randomly and independently selected in order to explore the parameter space (we choose 100), and the estimation algorithm is performed for a relatively small number of iterations under each set of initial values (we choose 10). We evaluate changes in the log-likelihood and select the set of initial values with the highest log-likelihood out of the 100 sets. For POCM, we learn the basis models using Algorithm 1. For HMM, we run the Baum-Welch algorithm for each group (Appendix B.4).

At the end of stage 1, we use the population average treatment effect ρ factor (Appendix F) to obtain the Treatment-II basis models for POCM, and subgroup HMMs under Treatment-II. Next, these models are used to initialize individual patient models for testing patients in stage 2. The treatment effect factor ρ can be estimated from comparative effectiveness trials with two treatment arms (Katzelnick et al., 2000). Note that the ρ is a simple way to generate a treatment effect in the simulated case study and not a necessary assumption in the POCM framework. If training data is also available for Treatment-II, stage 1 analysis can be done separately with observational data for patients on Treatment-II.

In stage 2, testing patients' PHQ-9 scores are generated for 6 periods under Treatment-I and 6 periods under Treatment-II from the ground truth (Appendix E.1). This process is similar to the STAR*D Study, in which 2876 patients with major depression underwent four levels of treatments involving adding and/or switching various antidepressant medications, psychotherapy, and other mood stabilizers depending on their symptom-free and treatment resistant status (Rush et al., 2006). Since the majority of patients in STAR*D only experienced two levels of treatment between 14 weeks to 12 months, our simulation setting involving two treatments with 6 months trials is practical for patients with persistent depression. We then estimate the testing patients' membership by performing Step 2 of the POCM algorithm (i.e., update the membership with fixed stage 1 basis models). Combining the membership and the basis models, we obtain their individual disease model $\hat{\boldsymbol{\pi}}_i, \hat{\mathbf{A}}_i, \hat{\mathbf{B}}_i$, $i = 1, \ldots, N^{\text{test}}$. The initial value of the membership is defined to be proportional to the inverse of the distance of the testing patient's profile to each cluster center of the training patients' profiles found in stage 1. In the HMM estimation, for each testing patient, we pick the subgroup model with the center closest to the patient profile as the initial values and then perform the Baum-Welch algorithm on their observations to obtain the personalized disease model.

We evaluate the performance of POCM vs. HMM in estimating the individual disease model using the average population difference between the true transition/emission matrices and the estimated matrices. The estimation error for a single patient is measured by the Frobenius norm (Horn & Johnson, 1990) defined as

$$\delta_i^{\mathbf{A}} = \sqrt{\sum_{s=1}^{|\mathcal{S}|}\sum_{s'=1}^{|\mathcal{S}|}\left[(\mathbf{A}_i^g - \hat{\mathbf{A}}_i)_{s,s'}\right]^2}, i = 1, \ldots, N. \tag{10}$$

where $\mathbf{A}_i^g$ is the ground-truth transition matrix for patient i, and $\hat{\mathbf{A}}_i$ is the estimated individual transition matrix. The maximum error is $\sqrt{2|\mathcal{S}|}$. We note in stage 1 of HMM, $\hat{\mathbf{A}}_i$ is the group model assigned to patient i. The population average estimation error for transition matrix is

$$\bar{\delta}^{\mathbf{A}} = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{\mathbf{A}}, \tag{11}$$

where $n = N^{\text{train}}$ or N^{test} for the learning stage or the fine-tuning stage respectively. Particularly, $\bar{\delta}^{\mathbf{A}}/|\mathcal{S}|$ is the average estimation error per element used to compare performance among models with different number of states. We define $\bar{\delta}^{\mathbf{B}}$ for the emission matrix in the same way.

4.2. Simulation of the decision stage: 12 treatment policies

In the decision stage, we compare the outcomes of 12 treatment policies for 24 periods. At each period, a treatment type is selected for each testing patient using one of the 12 policies in Table 2. We evaluate the performance gap between the POMDP based policies and other heuristic policies such as using Bayesian belief update only, PHQ-9 observations only, or fixed single treatment. These heuristics are included due to their ease of implementation in clinical practice.

Each policy is applied to the 200 testing patients, and each unique patient is simulated with 1000 repetitions. During each repetition, a patient's initial health state is drawn from the initial distribution estimated from stage 2. The outcome criterion is the average discounted total reward $\frac{1}{N}\sum_{t=1}^{T}\sum_{i=1}^{N}\phi^t r(s_{i,t}, a_{i,t})$, where N

Table 2
12 policies to be examined in the decision stage.

	Short name	Type	Description
1	`pomdp_true`	POMDP policy	Use the ground-truth individual parameters
2	`pomdp_basis`	POMDP policy	Use the group HMM model estimated in stage 1 that the patient is most likely belonging to.
3	`pomdp_pocm`	POMDP policy	Use the individual POCM estimated in stage 2
4	`pomdp_hmm`	POMDP policy	Use the individual HMM estimated in stage 2
5	`s_aggr`	Bayesian policy	Update the belief state by the Bayesian rule in Eq. (6), and select Treatment-II when $b(\mathrm{S}) \geq 0.2$, more aggressive
6	`s_cons`	Bayesian policy	Select Treatment-II when $b(\mathrm{S}) \geq 0.5$, more conservative
7	`h_aggr`	Bayesian policy	Select Treatment-II when $b(\mathrm{H}) \leq 0.8$, more aggressive
8	`h_cons`	Bayesian policy	Select Treatment-II when $b(\mathrm{H}) \leq 0.5$, more conservative
9	`o_aggr`	Observation policy	Select Treatment-II when the PHQ-9 is greater than 5, more aggressive
10	`o_cons`	Observation policy	Select Treatment-II when the PHQ-9 is greater than 10, more conservative
11	`tx_i_only`	Single treatment policy	Treatment-I for all periods
12	`tx_ii_only`	Single treatment policy	Treatment-II for all periods

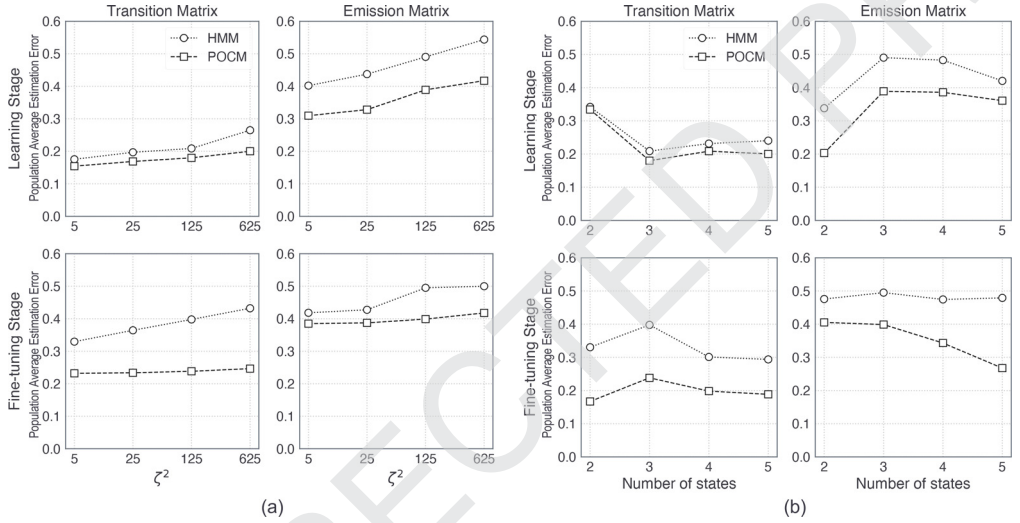

Fig. 3. Sensitivity analysis on the converged estimation error (Eq. (11)) between the POCM and the HMM. (a) Effect of latent structure significance, $\zeta^2 = 5, 25, 125, 625$; larger value of ζ^2 indicates strong latent structure. (b) Effect of number of health states, including 2, 3, 4, 5; Note that we divide the estimation error by the number of states to compare across models with different number of states.

is the number of patients and T is the number of treatment periods (N=200, T=24).

We test the policy performance on various parameter settings, including 3 levels of treatment effects ($\rho \in \{0.2, 0.5, 0.8\}$), 10 health utility structure ($\kappa(\mathrm{H}) \in \{0.8, 1\}$, $\kappa(\mathrm{M}) \in \{0.4, 0.8\}$, $\kappa(\mathrm{S}) \in \{0.1, 0.4\}$), and 9 cost structures ($h(\mathrm{I}) \in \{500, 1000\}, h(\mathrm{II}) \in \{1200, 2800\}$), totalling 270 scenarios (Appendix G.3). We perform sensitivity analyses to illustrate policy rankings under parameters uncertainty.

4.3. Numerical results

4.3.1. Parameter learning

We compared the performance of the POCM parameter inference and HMM inference using the population average estimation error defined in Eq. (11) in both the learning and the fine-tuning stages. The parameter μ was calibrated to a value of 0.1 to achieve the maximum objective (Appendix G.1). Results were based on various levels of significance of latent structure and the number of health states (Fig. 3). We can see that POCM performed better for both transition matrix and emission matrix estimation in the learning and the fine-tuning stages. This conclusion is not affected by the level of significance of latent structure or the number of health states. POCM has a fundamentally different structure about the patients' transition matrix that is composed of the previously learned basis models at the population level in stage 1, which makes the personalization in stage 2 less prone to being affected by observation uncertainty, bias, and outliers, and thus could be more clinically relevant.

In stages 1 and 2, there were NT observations under one treatment. In stage 1, a total of $2K(|\mathcal{S}|-1)|\mathcal{S}| + (K-1)K + N^{\text{train}}(K-1) = 2\times3\times2\times3 + 2\times3 + 800\times2 = 1,624$ free parameters were estimated in the POCM. $2K(|\mathcal{S}|-1)|\mathcal{S}| = 2\times3\times2\times3 = 36$ free parameters were estimated in the HMM. In stage 2, a total of $N^{\text{test}}(K-1) = 200\times2 = 400$ free parameters were estimated in the POCM for each treatment. A total of $2N^{\text{test}}(|\mathcal{S}|-1)|\mathcal{S}| = 2\times200\times2\times3 = 2400$ free parameters were estimated in the HMM for each treatment. On a MacBook with 2.2 GHz Quad-Core Intel Core i7 CPU, the computation time for POCM was 15.5 seconds per iteration in stage 1 and 0.25 seconds per patient in stage 2. The computation time for HMM was 0.55 seconds per iteration in stage 1 and 0.53 seconds per patient in stage 2. In summary, POCM took much longer in the learning stage (30 times of HMM considering an average of 100–200 iterations to converge), but it was slightly faster in the fine-tuning stage compared to HMM.

We tested several lengths of the fine-tuning stage, including 5, 10, 20, 50, and 100 periods. We found that POCM had a lower estimation error on transition and emission matrices than HMM on all testing lengths. The estimation error for the transition matrix of POCM decreased when the testing length increased from 5 to 20, and then stayed relatively flat when the testing length increased from 20 to 100. The estimation error for the emission matrix did not change significantly on different testing lengths (Appendix G.2).

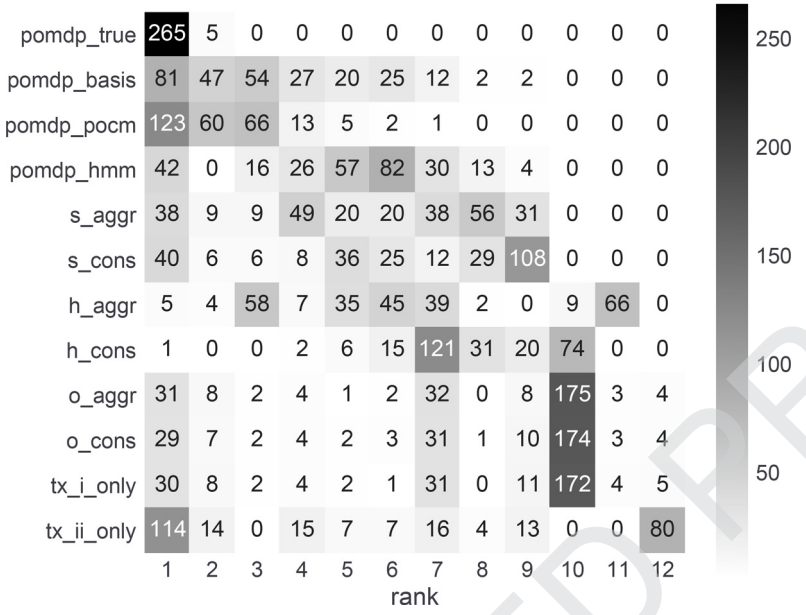

Fig. 4. Performance of the 12 treatment policies in 270 model settings. The numbers with rank 1–12 in all scenarios are displayed (including ties). Magnitude of the number is shown in greyscale.

4.3.2. Treatment outcomes and sensitivity analysis

We performed the policy comparison for 270 scenarios of parameter settings. In each scenario, we ranked the policies by their NMBs. We considered a tie when two or more policies' NMBs have less than a 1% difference (Appendix G.3). Fig. 4 lists the number of ranks each policy achieved among the 270 scenarios. We confirmed that `pomdp_true` achieved the most top ranks (5 out of 270 scenarios where it achieved rank 2 due to stochasticity in the simulation.) Policy `pomdp_pocm` had the second best performance with rank 1 to 3 in most scenarios. Policy `pomdp_basis`'s overall performance ranked the third among the POMDP policies, with fewer rank 1 than `pomdp_pocm`. Policy `pomdp_hmm` had the lowest performance among the POMDP policies. The average NMB difference between `pomdp_pocm` and `pomdp_basis` per scenario across the 270 scenarios is $13k. Therefore, `pomdp_pocm` is a better model to describe disease progression in a heterogeneous population. It outperformed `pomdp_basis` due to the fine-tuning of personalized treatment response model in stage 2, and outperformed `pomdp_hmm` because the HMM may overfit a model to an individual patient's limited observations at stage 2 (possibly with outliers), and it may also start with a worse initialization than POCM.

Furthermore, the heuristic policies performed worse than the POMDP policies; within each pair of the same type of heuristic policies (s, h, o), the aggressive policy performed better. In addition, we conducted an ordinal regression analysis examining what factors drive the performance difference between the POMDP policies (Appendix G.4). We examined factors including treatment effect, utility gap between health states, and treatment cost ratio (Appendix Table G.2). For `pomdp_pocm`, better ranks were associated with a smaller utility gap between depression states.

We also investigated the policy performance difference by subgroup of the testing patients (Appendix G.5). We divided 200 testing patients into four groups by their ground truth memberships. The first three groups were those with a membership close to 1 on one of the basis models: high-risk (51 patients), low-risk (53 patients), and stable (35 patients). The fourth group contained patients with no extreme membership on any basis model (62 patients). Appendix Figure G.4 shows the performance outcomes by subgroup under one parameter setting, including the group-averaged number of treatment switches, the proportion of Treatment-II assigned during stage 3, and the total reward in NMBs. We observed that the stable group had fewer treatment switches, fewer Treatment-II assignments, and higher rewards under most of the policies. On the contrary, the high-risk group had more frequent treatment switches, higher Treatment-II assignments, and lower rewards.

Lastly, we assumed that all patients shared the same Treatment-II effect ρ in all previous analyses; we relaxed this assumption by assigning different ρ's to the subgroups to reflect patients' diverse response to Treatment-II by their progression patterns. In this sensitivity analysis (Appendix G.6), we assigned the ground-truth $\rho = 0.2$ (very effective) to the high-risk subgroup, $\rho = 0.5$ (less effective) to the low-risk subgroup, and $\rho = 0.8$ (not effective) to the stable subgroup. In stage 2, we estimated the subgroup label of each testing patient using clustering. In stage 3, we collected the reward rankings of 12 treatment policies under 90 parameter settings. Results showed that `pomdp_pocm` is the best policy among all subgroups (except for `pomdp_true`), and it performed better for patients matched with the correct ground-truth ρ in stage 2 than those that were mismatched (Appendix Figure G.5). This result confirmed that POCM could be a better model than HMM for a heterogeneous population with diverse treatment response.

5. Conclusions and future work

In this paper, we proposed a three-stage POCM framework to estimate patient-specific chronic disease progression models for optimal treatment selection in a heterogeneous population. The POCM method makes the following assumptions: (1) The patient's health state is not fully observable; (2) The disease progression is a Markovian process; (3) There is a set of basis models, each representing a unique pattern of disease progression, and each patient's disease progression is a combination of these patterns; (4) Treatment can be tailored to individual patient online by learning his/her personal disease progression model using imperfect observations. We developed an efficient computational algorithm to estimate parameters of the POCM.

We designed a simulation case study of chronic depression treatment to demonstrate that the proposed POCM method can perform better than the standard estimation method, the HMM, in estimating individual disease models. In addition, we evaluated the performance of several adaptive treatment policies (POMDP policies and Bayesian policies) and simple heuristic policies based on past observations or fixed treatments. After applying all 12 policies to the 200 testing patients in the decision stage, we assessed their performance in NMBs and reached three main conclusions: (1) The POCM achieved lower error than HMM in estimating a personalized disease model; (2) The POCM-based POMDP policy gave the highest reward under most settings of treatment effect, utility structure, and cost structure. In particular, among the POMDP policies, POCM-based policy had better performance than the HMM-based policies; (3) In most cases, treatment policies performed similarly across subgroups. The high-risk group had more treatment switches, more Treatment-II selection, and lower rewards under most policies.

The proposed POCM is one method to estimate a unique personalized model in the broader context of precision medicine, in which we hope such a model can assist clinicians' own expertise and judgment in selecting treatment. Current debate on using computer algorithms to automatically make clinical decisions centers on the continuum between "fully human-guided vs fully machine-guided data analysis" (Beam & Kohane, 2018), and there are some fears in the psychiatry community that "reliance on big data to inform treatment decisions might lead to ignoring experiences and values... Computer generated recommendations may carry a false authority that would override expert human judgment" (Simon & Yarborough, 2020). However, even advanced AI models require human inputs that use domain knowledge to clean the dataset, define features and/or parameters, and set the objective of the optimization. To validate the proposed method in practice, we should first evaluate the estimated population basis matrices against known medical knowledge on the disease's natural history. In the long term, the best way to validate the proposed method is to conduct a human-subject study. In the case of chronic depression, all patients are monitored with a gold-standard test and the PHQ-9 to assess their depression severity in each period (this effort can be aided with personal sensing technology). Using these data, we can compare the estimated true progression model from the gold-standard test with the estimated progression modeled by POCM using PHQ-9 scores as observations. In addition, we can conduct a comparative effectiveness trial, where the control arm is the usual care, and the treatment arm is the POCM framework. Outcomes can include NMBs and depression-free days (remission periods). In fact, our simulation experiment is similar to conducting a hypothetical 12-armed trial.

This study has several limitations in both methodology and practice. First, we excluded death in the POMDP. The effect of depression treatment on suicide rate is controversial; some studies showed an increase in the suicide rate among teens, while other studies showed no effects. Overall, the suicide rate is very low among chronically depressed patients (Simon et al., 2016). If the death rate differs significantly by treatment in another application, the transition matrix estimation can be adjusted to include a death state in the learning stage. Second, accurate estimation of a person-level emission matrix is challenging. The PHQ-9 is a validated test with a sensitivity and specificity of 88% (Kroenke & Spitzer, 2002) which can be used to initialize the emission matrix. There are multiple reasons to estimate an individualized emission matrix. Similar to screening for risk factors (e.g., risky sexual behavior, drug use) of some sexually transmitted diseases that are stigmatized in society, patients may be less willing to reveal the truth about mental illness. This willingness to be truthful may vary among persons. One way to address this estimation problem is to conduct an observational study that collects depression progression information using a gold-standard test, the PHQ-9, and demographic and clinical profiles. Using these data, a regression model can be built to initialize a personal emission matrix. Third, we only demonstrated the performance of POCM by simulating two treatments in the depression case study. The three-stage framework can be extended to applications with more than two treatments. In stage 1, we can either estimate ρ_1 to ρ_m from clinical trials (where m is the treatment index), or use separate EHR dataset under each treatment to directly estimate the basis models. In stage 2, testing patients go through m separate trials to estimate a personal disease model under each treatment. Then in stage 3, the actions include selecting treatment 1 to treatment m. Higher number of treatment increases data requirement in stage 1 and longer duration in stage 2, which brings additional practical challenges. Fourth, the three-stage framework can be seen as too rigid or impractical. It is possible to combine or repeat stage 2 and stage 3 with continuous online learning during treatment selection or skip stage 2 all together if individual variations within a subgroup are small. For example, the combined stages can be similar to an adaptive sequential trial where treatment modification is done at predefined time points (Bothwell, Avorn, Khan, & Kesselheim, 2018; Chow, 2014).

The POCM modeling framework does not apply broadly in the following cases. The model assumes the underlying health state transition process is an HMM, but not all longitudinal clinical data meet this assumption. For example, some longitudinal disease trajectories are transient processes and never reach steady state. If the decision problem has high-dimensional state space and action space, fine-tuning may not be an efficient estimation procedure for both HMM and POCM. Furthermore, the basis models should represent canonical patterns in the target population. If data from existing population do not represent broad patterns in future individuals, then POCM is not suited to model these conditions.

There are several directions to expand this research. First, although we only presented one disease application in chronic depression, POCM can be applied to a broader range of chronic diseases that meet similar assumptions on partial-observable health state, the disease process is Markovian, and long treatment duration with treatment switching options. In addition, POCM is not limited to the medical decision-making problem. Take machine maintenance as an example, the state of the machine can change over time, and the probability of state transition is different for each individual machine. Therefore, the health progression of an individual machine can be modeled with POCM, and machine maintenance policies can be tailored to an individual machine by using the basis models and membership learned by POCM. Another example is personalized health management from daily behavioral data (Xiao, 2017, chap. 5). The fast-growing development of sensing devices enable the continuous monitoring of human behavior (such as physical activity and food intake), and health state measurements such as the body mass index (BMI). A personalized health management program such as obesity prevention can be achieved by learning the basis behavior models with POCM, which lead to individual behavior model, and then we can find an optimal plan of health activity via POMDP.

In summary, we developed a three-stage POCM framework to estimate a personal model of chronic disease progression using both population data and treatment experimentation to optimally select treatment. We designed a simulation case study on chronic depression. Results showed that the POCM framework can lead to better performance on individual parameter estimation over the traditional HMM method. This framework is promising for modeling the chronic disease progression process and developing a personalized adaptive treatment plan for individual patients in a heterogeneous population.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2023.03.014.

References

Ayer, T. (2015). Inverse optimization for assessing emerging technologies in breast cancer screening. *Annals of Operations Research, 230*(1), 57–85.
Ayer, T., Alagoz, O., & Stout, N. K. (2012). OR Forum—a POMDP approach to personalize mammography screening decisions. *Operations Research, 60*(5), 1019–1034.
Ayer, T., Zhang, C., Bonifonte, A., Spaulding, A. C., & Chhatwal, J. (2019). Prioritizing hepatitis C treatment in US prisons. *Operations Research, 67*(3), 853–873.
Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics, 41*(1), 164–171.
Beam, A., & Kohane, I. (2018). Big data and machine learning in health care. *JAMA, 319*(13), 1317–1318.
Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer-Verlag.
Bothwell, L. E., Avorn, J., Khan, N. F., & Kesselheim, A. S. (2018). Adaptive design clinical trials: A review of the literature and clinicaltrials.gov. *BMJ Open, 8*(2). https://doi.org/10.1136/bmjopen-2017-018320. https://bmjopen.bmj.com/content/8/2/e018320
Calandriello, D., Lazaric, A., & Restelli, M. (2014). Sparse multi-task reinforcement learning. In *Advances in neural information processing systems* (pp. 819–827).
Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D., & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal, 321*(7262), 694.
Cassandra, A., Littman, M. L., & Zhang, N. L. (1997). Incremental pruning: A simple, fast, exact method for partially observable Markov decision processes. In *Proc. 13th conf. on uncertainty in artificial intelligence* (pp. 54–61). Morgan Kaufmann Publishers Inc..
Chen, Q., Ayer, T., & Chhatwal, J. (2018). Optimal M-switch surveillance policies for liver cancer in a hepatitis C–infected population. *Operations Research, 66*(3), 673–696.
Chow, S.-C. (2014). Adaptive clinical trial design. *Annual Review of Medicine, 65*(1), 405–415. https://doi.org/10.1146/annurev-med-092012-112310. PMID: 24422576
D'Alfonso, S. (2020). Ai in mental health. *Current Opinion in Psychology, 36*, 112–117.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1–22.
Dickstein, K., Cohen-Solal, A., Filippatos, G., McMurray, J. J., Ponikowski, P., Poole-Wilson, P. A., … Hoes, A. W. (2008). ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008. *European Journal of Heart Failure, 10*(10), 933–989.
Faissol, D. M., Griffin, P. M., & Swann, J. L. (2007). Timing of testing and treatment of hepatitis C and other diseases. In *Proceedings* (p. 11).
Guideline Development Panel for the Treatment of Depressive Disorder (2019). Clinical practice guideline for the treatment of depression across three age cohorts. https://www.apa.org/depression-guideline/. Accessed: 2020-06-19.
Helm, J. E., Lavieri, M. S., Van Oyen, M. P., Stein, J. D., & Musch, D. C. (2015). Dynamic forecasting and control algorithms of glaucoma progression for clinician decision support. *Operations Research, 63*(5), 979–999.
Horn, R. A., & Johnson, C. R. (1990). *Matrix analysis*. Cambridge University Press.
Jaakkola, T., Singh, S. P., & Jordan, M. I. (1995). Reinforcement learning algorithm for partially observable Markov decision problems. In *Advances in neural information processing systems* (pp. 345–352).
Jameson, J., & Longo, D. (2015). Precision medicine–personalized, problematic, and promising. *The New England Journal of Medicine, 372*(23), 2229–2234.
Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., … Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*(4), 230–243.
Katzelnick, D. J., Simon, G. E., Pearson, S. D., Manning, W. G., Helstad, C. P., Henk, H. J., … Kobak, K. A. (2000). Randomized trial of a depression management program in high utilizers of medical care. *Archives of Family Medicine, 9*(4), 345–351.
Kazemian, P., Helm, J. E., Lavieri, M. S., Stein, J. D., & Van Oyen, M. P. (2019). Dynamic monitoring and control of irreversible chronic diseases with application to glaucoma. *Production and Operations Management, 28*(5), 1082–1107.
Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals, 32*(9), 509–515.
Lavieri, M. S., Puterman, M. L., Tyldesley, S., & Morris, W. J. (2012). When to treat prostate cancer patients based on their PSA dynamics. *IIE Transactions on Healthcare Systems Engineering, 2*(1), 62–77.
Lee, E., Lavieri, M. S., Volk, M. L., & Xu, Y. (2015). Applying reinforcement learning techniques to detect hepatocellular carcinoma under limited screening capacity. *Health Care Management Science, 18*(3), 363–375.
Li, H., Liao, X., & Carin, L. (2009). Multi-task reinforcement learning in partially observable stochastic environments. *Journal of Machine Learning Research: JMLR, 10*, 1131–1186. http://dl.acm.org/citation.cfm?id=1577069.1577109
Lin, Y., Huang, S., Simon, G. E., & Liu, S. (2016). Analysis of depression trajectory patterns using collaborative learning. *Mathematical Biosciences, 282*, 191–203.
Lin, Y., Liu, K., Byon, E., Qian, X., Liu, S., & Huang, S. (2017). A collaborative learning framework for estimating many individualized regression models in a heterogeneous population. *IEEE Transactions on Reliability, 67*(1), 328–341.
Lin, Y., Liu, S., & Huang, S. (2018). Selective sensing of a heterogeneous population of units with dynamic health conditions. *IISE Transactions, 50*(12), 1076–1088.
Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*(2), 129–137.
Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 44*(2), 226–233.
Madadi, M., Zhang, S., & Henderson, L. M. (2015). Evaluation of breast cancer mammography screening policies considering adherence behavior. *European Journal of Operational Research, 247*(2), 630–640. https://doi.org/10.1016/j.ejor.2015.05.068. https://www.sciencedirect.com/science/article/pii/S0377221715004804
Maillart, L. M., Ivy, J. S., Ransom, S., & Diehl, K. (2008). Assessing dynamic breast cancer screening policies. *Operations Research, 56*(6), 1411–1427.
Mason, J. E., England, D. A., Denton, B. T., Smith, S. A., Kurt, M., & Shah, N. D. (2012). Optimizing statin treatment decisions for diabetes patients in the presence of uncertain future adherence. *Medical Decision Making, 32*(1), 154–166.
Musliner, K. L., Munk-Olsen, T., Eaton, W. W., & Zandi, P. P. (2016). Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes. *Journal of Affective Disorders, 192*, 199–211.
Negoescu, D. M., Bimpikis, K., Brandeau, M. L., & Iancu, D. A. (2017). Dynamic learning of patient response types: An application to treating chronic diseases. *Management Science, 64*(8), 3469–3488.
Neumann, P. J., Cohen, J. T., & Weinstein, M. C. (2014). Updating cost-effectiveness—the curious resilience of the $50,000 per QALY threshold. *The New England Journal of Medicine, 371*(9), 796–797.
Otten, M., Timmer, J., & Witteveen, A. (2020). Stratified breast cancer follow-up using a continuous state partially observable Markov decision process. *European Journal of Operational Research, 281*(2), 464–474. https://doi.org/10.1016/j.ejor.2019.08.046. https://www.sciencedirect.com/science/article/pii/S0377221719307167
Phillips, K. A. (2018). Precision medicine: From science to value. *Health affairs (Project Hope), 37*(5), 694–701.
Pratt, L., & Brody, D. (2014). Depression in the us household population, 2009–2012. NCHS Data Brief, Hyattsville, MD.
Ross, E. L., Vijan, S., Miller, E. M., Valenstein, M., & Zivin, K. (2019). The cost-effectiveness of cognitive behavioral therapy versus second-generation antidepressants for initial treatment of major depressive disorder in the united states: A decision analytic model. *Annals of Internal Medicine, 171*(11), 785–795. https://doi.org/10.7326/M18-1480.
Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., … Lebowitz, B. D., et al., (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *The American Journal of Psychiatry, 163*(11), 1905–1917.
Saure, A., Patrick, J., Tyldesley, S., & Puterman, M. L. (2012). Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research, 223*(2), 573–584.
Schulam, P., & Saria, S. (2015). A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure. In *Advances in neural information processing systems* (pp. 748–756).
Shechter, S., Bailey, M., Schaefer, A., & Roberts, M. (2008). The optimal time to initiate HIV therapy under ordered health states. *Operations Research, 56*(1), 20–33.
Simon, G. E., Coleman, K. J., Rossom, R. C., Beck, A., Oliver, M., Johnson, E., … Shortreed, S. M., et al., (2016). Risk of suicide attempt and suicide death following completion of the patient health questionnaire depression module in community practice. *The Journal of Clinical Psychiatry, 77*(2), 221.
Simon, G. E., & Yarborough, B. J. (2020). Good news: Artificial intelligence in psychiatry is actually neither. *Psychiatric Services, 71*(3), 219–220.
Skandari, M. R., & Shechter, S. M. (2021). Patient-type bayes-adaptive treatment plans. *Operations research, 69*(2), 574–598. https://doi.org/10.1287/opre.2020.2011.
Tanaka, F., & Yamamura, M. (2003). *Multitask reinforcement learning on the distribution of MDPs: vol. 123* (pp. 1108–1113). https://doi.org/10.1109/CIRA.2003.1222152.
Twisk, J., & Hoekstra, T. (2012). Classifying developmental trajectories over time should be done with great caution: A comparison between methods. *Journal of Clinical Epidemiology, 65*(10), 1078–1087.
Wang, X., Sontag, D., & Wang, F. (2014). Unsupervised learning of disease progression models. In *Proc. 20th ACM SIGKDD internat. conf. on knowledge discovery and data mining* (pp. 85–94). ACM.
Xiao, C. (2017). *Optimization and machine learning methods for medical and healthcare applications*. University of Washington Ph.D. thesis.
You, M., Liu, B., Byon, E., Huang, S., & Jin, J. J. (2018). Direction-dependent power curve modeling for multiple interacting wind turbines. *IEEE Transactions on Power Systems, 33*(2), 1725–1733.
Zhu, P., Li, X., Poupart, P., & Miao, G. (2018). On improving deep reinforcement learning for POMDPs. arXiv preprint arXiv:1804.06309.