2

How AI Can Help Depression Treatment

Designing Patient-Specific Adaptive Interventions

Shan Liu and Shuai Huang

Summary of the Problem

Major depressive disorder, one of the most common mental disorders in the United States, if left untreated in time can lead to disability, reduced quality of life and productivity, and increased risk of death due to comorbid conditions and suicide. Inadequate follow-up care is a major shortcoming in current depression treatment that can lead to poor quality of care and high cost.

Summary of the Solution

Designing artificial intelligence (AI)-assisted technology to better understand the disease trajectory and further develop appropriate strategies for monitoring and treatment of major depression under resource constraints is an important and challenging task. In this chapter, we present seven studies that developed methods for AI-assisted, data-driven decision support systems to aid healthcare professionals. These methods focus on modeling chronic depression's complex disease trajectories, identifying patients at high risk of progression, and recommending adaptive and cost-effective follow-up care. Long-term goals of this research include improving patient health outcomes and facilitating efficient allocation of healthcare providers' limited resources through the use of novel technology.

Summary of Relevance in a Post-COVID World

The economic losses and isolation resulting from non-pharmaceutical interventions deployed to slow the spread of COVID-19 have exacerbated the

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

challenges of mental health and suicide. They have simultaneously increased the risk of in-person counseling and the availability of such appointments. This increased demand and reduced supply for depression treatment make more important than ever the use of technology driven methods to optimize the deployment of counselling resources.

2.1 Chronic Depression as a Significant Public Health Problem

Depression is a complex, dynamic mental disorder characterized by sad mood, loss of interest in activities, weight gain or loss, psychomotor agitation or retardation, fatigue, inappropriate guilt, difficulties concentrating, and recurrent thoughts of death [1]. Depression is diagnosed by five or more of the foregoing symptoms presenting for a continuous period of at least two weeks. It is one of the most common mental disorders in the United States, affecting more than 10% of the population [2]. Undiagnosed depression can lead to disability, reduced quality of life, reduced productivity, and increased risk of death due to comorbid chronic conditions and suicide [3, 4]. Though remarkable progress has been made in reducing the mortality and morbidity burdens for many diseases, including stroke, heart disease, and HIV/AIDS, depression-related morbidity and mortality has been rising in recent decades [5]. Suicide has recently become the number 1 cause of violent death and the 10th leading cause of death in the United States [6, 7]. National and state governments as well as guideline-setting bodies are making efforts to address this urgent problem. The U.S. Preventive Services Task Force (USPSTF) updated depression screening guidelines in 2016 that recommend "screening for depression in the general adult population," and "screening should be implemented with adequate systems in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up" [8, 9]. In addition, the concurrence of depression with other chronic conditions and substance addiction is an understudied area where patient outcomes are known to be poor [10]. An open challenge is to design effective monitoring and treatment strategies for depression and its associated comorbid conditions.

Treatment for depression includes psychotherapy, antidepressants, or a combination of the two with supportive care. Due to potential side effects of the medications, the Food and Drug Administration (FDA) emphasizes that patients taking antidepressants should be closely monitored [3]. Finding appropriate strategies for routine monitoring is important and can be controversial [11].

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

Current recommendations for follow-up care of chronic depression are based almost entirely on expert opinion, which could suggest semiannual or annual monitoring intervals [12]. These recommendations do not account for significant heterogeneity in the course of depression between individuals and within individuals over time. Given that as many as 30 million Americans use antidepressants [2, 13], even minor changes in recommendations for follow-up frequency have major implications for healthcare utilization and cost.

In recent years, there has been an explosion of healthcare technology development including mobile health apps and telehealth that aim to merge big data analytics and AI to support healthcare services. Foreseeable benefits of AI-assisted decision support systems include faster, safer, cheaper, more convenient, and higher quality of care tailored to individual patients. There is strong interest from the private sector to commercialize remote monitoring and treatment platforms for chronic depression. In this chapter, we first briefly discuss the methodological challenges and relevant literature in Section 2.2, and then show highlights our recent work in developing AI-based methods for chronic depression in Section 2.3.

2.2 Methodological Challenges to Optimize Care for Chronic Depression

The ultimate goal for AI-assisted technology in the care of chronic depression is to help mental healthcare providers create patient-specific monitoring and treatment strategies that lower the risk of future recurrence of depression symptoms. This technology aims to enable healthcare systems to efficiently identify patients who would benefit from proactive management of their symptoms as well as support targeted performance metrics to evaluate the success of clinical interventions. Development of such technology requires a systems perspective and an associated computational platform, and a seamless integration with decision-analytic methods to assess their cost-effectiveness.

There are three major methodological challenges: (i) effectively analyze heterogeneous depression trajectories of a patient population and proactively probe new trajectories, (ii) adequately characterize the disease progression processes that govern these trajectories and design adaptive interventions, and (iii) design rigorous cost-effectiveness analyses to evaluate the cost and benefit of the proposed technology across subgroups of patients. To address these challenges, we conducted a literature search and identified the following research gaps.

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

2.2.1 Research Gaps

2.2.1.1 Gap 1: Discovering Depression Progression Patterns

Disease progression is often modeled mathematically using data that can help quantify the dynamic and temporal relationships of outcome measurements. Statistical-based learning methods have a long history. Common techniques include regression, Bayesian updating, discrete-time Markov models, hidden Markov models, semi-Markov models, hidden semi-Markov models, continuous-time Markov models, Markov random fields, neural networks, and other supervised/unsupervised machine learning approaches. Applications of these methods can be seen in modeling CD4 count decline in HIV patients [14], liver deterioration in patients on the transplant waiting list [15–17], depression progression [18, 19], and chronic obstructive pulmonary disease progression [20]. However, the majority of these studies ignore individual heterogeneity and subgroup progression patterns in the disease process and instead only consider a single stochastic process meant to reflect average, population-level outcomes.

2.2.1.2 Gap 2: Designing Adaptive Healthcare Interventions

The literature on using stochastic and dynamic models to optimize disease screening and treatment decisions over time is extensive. For example, Markov Decision Processes (MDPs), dynamic programming, and reinforcement learning have been used to decide how to optimally monitor and control disease progression [21]. Despite the successful application of these methods in a number of health applications [14–17], they are often population-based, require extensive data to estimate the transition probabilities and rewards for each possible action, and often do not incorporate real-time updating of the disease process and model parameters using all available information.

2.2.1.3 Gap 3: Linking Adaptive Technology to Cost-Effective Clinical Management

Cost-effectiveness analysis (CEA) is a crucial methodological component when designing and evaluating new technologies to enable their adoption into routine clinical practice [22, 23]. Two of the most important questions regarding depression are whether routine monitoring is justified and how to switch treatment. These questions are complicated by the heterogeneity in disease progression and treatment response within the population. With recent advancements in mobile phone apps, remote sensing, telehealth platforms, and big data, there is a considerable interest in developing commercial applications of digital therapeutics and automated remote monitoring of depression. The cost-effectiveness of this technology is uncertain and should be carefully

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

modeled, accounting for population heterogeneity and differential treatment outcomes. A CEA of AI-assisted chronic depression management must link the data-driven design of adaptive interventions with evaluation of long-term patient health outcomes and costs in real-world implementation scenarios.

2.2.2 A New AI-Assisted Technology Framework

To remove these methodological barriers, the state-of-the-art statistical modeling, optimization, and decision-analytic modeling can help to create an AI-assisted technology framework for development and evaluation of adaptive chronic depression interventions. We proposed the following three steps to accomplish such a framework in our recent work:

Step 1: Discover patterns in chronic depression and suicide ideation trajectories using longitudinal person-level symptom severity measurements from electronic health record (EHR) data, and build models for depression progression dynamics using statistical learning methods.

Step 2: Create machine learning and optimization models to predict future risk of depression progression and treatment response. We discuss several online algorithms to conduct adaptive monitoring and treatment selection in Section 2.3.

Step 3: Evaluate the cost-effectiveness of adaptive depression interventions compared with current clinical practice and national guidelines on the population level. Outcome measures include monitoring accuracies (i.e., sensitivity and specificity of the monitoring technology), cost, life-years gained, quality-adjusted life years (QALYs) gained, and incremental cost-effectiveness ratios between strategies of interest.

Table 2.1 provides a comparison of the current depression care practice guidelines at a representative healthcare system and the proposed framework.

2.2.2.1 Guide to the Literature

For disease-trajectory modeling and depression trajectories in particular, we refer the readers to the following papers: Twisk and Hoekstra 2012 compared methods to classify developmental trajectories over time [24]; Craig and Sendi 2002 presented a tutorial on the estimation of transition matrix of a discrete-time Markov chain with healthcare examples [25]; and Musliner et al. 2015 conduced a systematic review of long-term trajectories of depressive symptoms [26]. For readers looking for technical materials on pattern recognition, classification, sequential data, neural network, and Markov models, we refer them to Bishop's textbook on pattern recognition and machine learning [27]. For readers interested in an introduction to the theory and practice in AI, we refer them to Russell and Norvig's textbook on AI [28]. For more information

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

Table 2.1. *Comparison of current practice and proposed framework*

	Current practice	Proposed framework
Objective	Passive information collection	Proactively prevent and treat a patient's chronic depression symptoms
Method	Routine clinic visits with fixed frequency	Enable AI-assisted interventions that are adaptive to individual disease trajectory
Capability	Assessment of depression severity and suicide risk	1) Quantify and predict individual patient's depression and suicide ideation trajectory and risk 2) Determine adaptive monitoring schedule, treatment selection, and proactively collect information
Cost-effectiveness	Not evaluated	Evaluated by decision-analytic models using simulation and scenario analyses

on how to conduct a cost-effectiveness analysis, we refer the readers to Drummond's textbook on economic evaluation in healthcare [29].

2.3 Highlights from Seven Research Studies

In this section, we present highlights from our recent work. In these studies, we are developing methods to enable the AI-assisted technology framework introduced in Section 2.2. With expanded use of EHR, many health systems can administer and store longitudinal depression measurements for a large number of patients. The majority of studies described in this section used data from the Mental Health Research Network (MHRN) – a consortium of research centers affiliated with 11 large health systems. The dataset contains depression screening and outcomes data with approximately 2 million observations from a diverse and representative sample of outpatients in 5 states (California, Colorado, Minnesota, Washington, and Idaho).

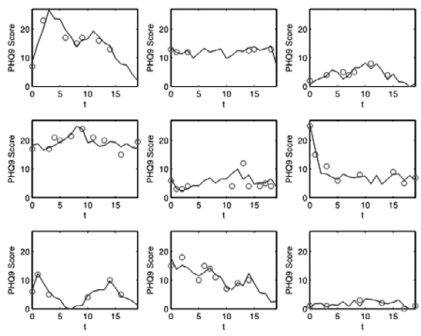
Our MHRN EHR dataset includes de-identified person-level Patient Health Questionnaire (PHQ)-9 depression measures (total scores and individual item scores) between year 2007 and 2012. PHQ-9 is a self-administrated questionnaire that includes 9 multiple-choice questions with a total score ranging from 0 to 27 [30]. PHQ-9 stratifies depression into 5 severity levels including no depression (0–4), mild depression (5–9), moderate depression (10–14), moderate severe depression (15–19), and severe depression (20–27). The data set also contains relative time between measurements; type of provider (primary care, specialist, and mental health); individuals' age, sex, race/ethnicity, diagnosis, and treatment status; and the Charlson Comorbidity Index score.

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

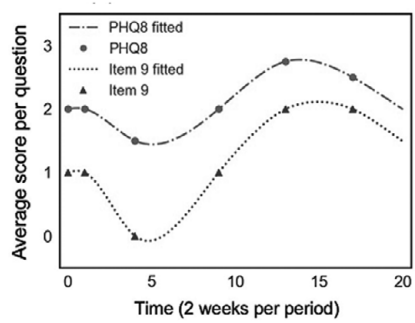
The Charlson Comorbidity Index score is a summary score of medical disease burden including a total of 22 conditions (each condition is assigned a score of 1, 2, 3, or 6, depending on the associated mortality risk) that aims to predict the 1-year mortality [31]. The majority of patients are older than age 45 (age 18–29, 13%; 30–44, 27%; 45–64, 43%; 65+, 17%) and female (70%).

2.3.1 Modeling Chronic Depression Trajectories

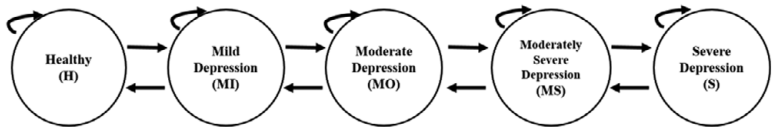
Depression trajectories are often estimated from noisy, sparse, and irregular person-level data. Since these time series do not follow any known functional forms, several alternative methods can be used to model them: (i) **Smoothing B-spline** (Figure 2.1a) can be used to characterize nonlinear patterns. An irregular observational time interval is first transformed to the B-splines bases, then the trajectory signals are represented as linear combinations of these bases and can be computed recursively for any desired degree of the polynomial using the algorithms in Boor [32]. (ii) **Gaussian process regression** (GPR) (Figure 2.1b) is a Bayesian nonparametric method that transforms observations to a longitudinal probability distribution [33]. The rational quadratic kernel can be used,



a) Lines are the smoothing B-spline fits on 9 randomly selected depression patients. Circles represent their observed PHQ-9 scores. X-axis unit: biweek.



b) Each line is the GPR model fit for one patient's average per-item scores (y-axis). X-axis unit: biweek.



c) A Markov model representation with five depression severity states.

Figure 2.1 Three alternative representations of chronic depression disease progression (panel b is from [35]).

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

and an optimal set of hyperparameters can be found that maximizes the marginal likelihood. (iii) **Discrete-time Markov model** is used with a state transition matrix that predicts the distribution of disease states over time (Figure 2.1c). The Expectation-Maximization (EM) algorithm can be applied to impute the missing data and obtain maximum likelihood estimators [25]. Alternatively, irregular individual observations can be first fitted with a smoothing spline and then partitioned into regular time intervals. Next, a transition matrix of movements between states is created by counting the number of transitions from each disease state to other states at each time interval [34]. The three alternative representations of the disease process may bring different advantages in the subsequent pattern discovery tasks shown in studies one, two, and four.

2.3.1.1 Pattern Discovery

We describe several studies that applied statistical learning and artificial neural networks to discover patterns in chronic depression progression. Two assumptions are made based on domain knowledge in chronic depression: (i) There are latent disease-trajectory patterns in the population (e.g., subgroups defined by stable mild, increasing severity, fluctuating severity). Patients in different subgroups may follow significantly different progression trajectories, but patients within the same subgroup may follow similar progression trajectories. (ii) Similarity information between patients can be quantified by comparing patients' demographic and clinical profiles from which features can be drawn to predict similarities in patients' future disease progressions.

Study One: Collaborative Modeling In Lin et al. 2016 [36], we analyzed patterns in the collected depression trajectories of a treatment population and compared several methods to predict individual trajectories for monitoring treatment outcomes. The data include longitudinal PHQ-9 scores over 4 years for assessing depression severity from the MHRN. We analyzed >3,000 patients with at least six PHQ-9 observations who received treatment longer than 6 months. We used smoothing splines to model individual depression trajectories. We then used K-means clustering and collaborative modeling (CM) to identify subgroup patterns. We found five broad trajectory patterns in the ongoing treatment population: stable high, stable low, fluctuating moderate, an increasing and a decreasing group (Figure 2.2a).

The CM approach assumes that the heterogeneous progression dynamics are represented by a number of canonical models in the population, where each patient's progression model is captured as variants of these canonical models [36–38]. CM considers the underlying cluster structure embedded in the

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

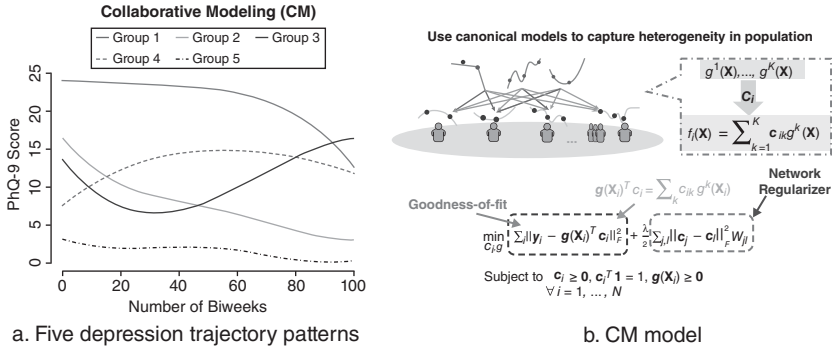


Figure 2.2 CM formulation and result [36]

population and the resemblance of the individuals to these clusters. We assign a membership vector, denoted as $\boldsymbol{c}_i = [c_{i1}, \ldots, c_{iK}]^T$, to each patient i, while K is the number of canonical pattern groups. Thus, c_{ik} denotes the probability of patient i belonging to a group k. For each patient i, we assume that there are longitudinal measurements (i.e., PHQ-9 scores) at n_i time points, denoted as $\boldsymbol{y}_i = \left[y_{i1}, \ldots, y_{in_i}\right]^{\mathrm{T}} \in \mathbb{R}^{n_i \times 1}$, and the longitudinal measurements of the risk factors, denoted as $\mathbf{X}_i = \left[\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\right]^{\mathrm{T}} \in \mathbb{R}^{n_i \times p}$. CM employs a canonical model $g^k(\mathbf{X})$ for characterizing the trajectory of group k. The canonical model is flexible and can take the form of a B-spline model, a GPR, or a Markov model. Then, the progression model of patient i can be characterized by a weighted combination of the canonical models of the K latent subgroups, $f^i(\boldsymbol{X}) = \sum_{\boldsymbol{k}} \boldsymbol{c_{ik}} \boldsymbol{g^k}(\boldsymbol{X})$. Furthermore, the similarity between two patients can be quantified by comparing their risk factors and past trajectory information. The similarity can be represented as a similarity matrix, $\mathbf{W}$, with each element, w_{jl}, representing the similarity between a pair of patients, j, l, and reflecting how likely the patients' progressions would be similar. To guide the learning of model parameters, we can minimize the least square loss function (Figure 2.2b) or maximize the log likelihood to measure the goodness-of-fit of the individual progression models.

Study Two: Artificial Neural Network Depression is often accompanied by thoughts of self-harm, which are a strong predictor of subsequent suicide attempt and suicide death. Few empirical data are available regarding the temporal correlation between depression symptoms and suicidal ideation. In Gong et al. 2019 [35], we investigated the traditional concern that suicidal ideation may increase during a period of depression improvement using

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

depression trajectory data. We analyzed a chronic depression treatment population's EHR, which contained 610 patients' longitudinal PHQ-9 scores within 20 two-week periods. We discovered patterns in trajectories of depressive symptoms using GPR and artificial neural networks. We also estimated correlations between symptomatology (PHQ-8) and suicide ideation (Item 9). We found five patterns in the PHQ-8 trajectories. PHQ-8 and Item 9 scores displayed strong temporal correlations. See Figure 2.3. We also found 8% to 13% of the patients have experienced an increase in suicidal ideation during improvement of their PHQ-8. We showed some evidence that subgroups of depressive patients are at increased risk of suicide ideation during PHQ-8 improvement.

2.3.2 Designing Adaptive Interventions

2.3.2.1 Adaptive Monitoring

Study Three: Rule-Based model In Lin et al. 2018 [39], we established a rule-based method to identify a set of risk predictive patterns from person-level longitudinal depression measurements by integrating three steps: data transformation, rule discovery, and rule evaluation. We further used the identified rules to create rule-based monitoring strategies to adaptively monitor patients. To evaluate the effectiveness of rule-based monitoring, we compared several monitoring strategies by estimating the number of depressive patients (PHQ-9 $\geq$ 10) in the next 6 months that are correctly monitored (true positives). We assumed under the status quo that all patients are monitored every 6 months, which may lead to unnecessary monitoring of low-risk patients. We also considered a PHQ-9-based strategy, which monitors the patient if his/her last-period PHQ-9 score is 10 or greater. Under rule-based monitoring, we considered both using individual rules and combining all top predictive rules.

We applied the rule-based method on the EHR data of a depression treatment population containing PHQ-9 scores. Twelve risk predictive rules were identified (Table 2.2). We found the rule-based prognostic model based on the identified rules enabled more accurate prediction of disease severity than other prognostic models, including RuleFit, logistic regression, and Support Vector Machine. Two rule-based monitoring strategies outperformed the latest PHQ-9-based monitoring strategy by providing higher sensitivity and specificity. We concluded that the rule-based method can lead to a better understanding of disease dynamics and achieve more accurate prognostics of disease progressions (Figure 2.4).

Study Four: Selective Sensing In Lin et al. 2018b [38], the study's objective was to build personalized trajectory models to proactively probe new

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

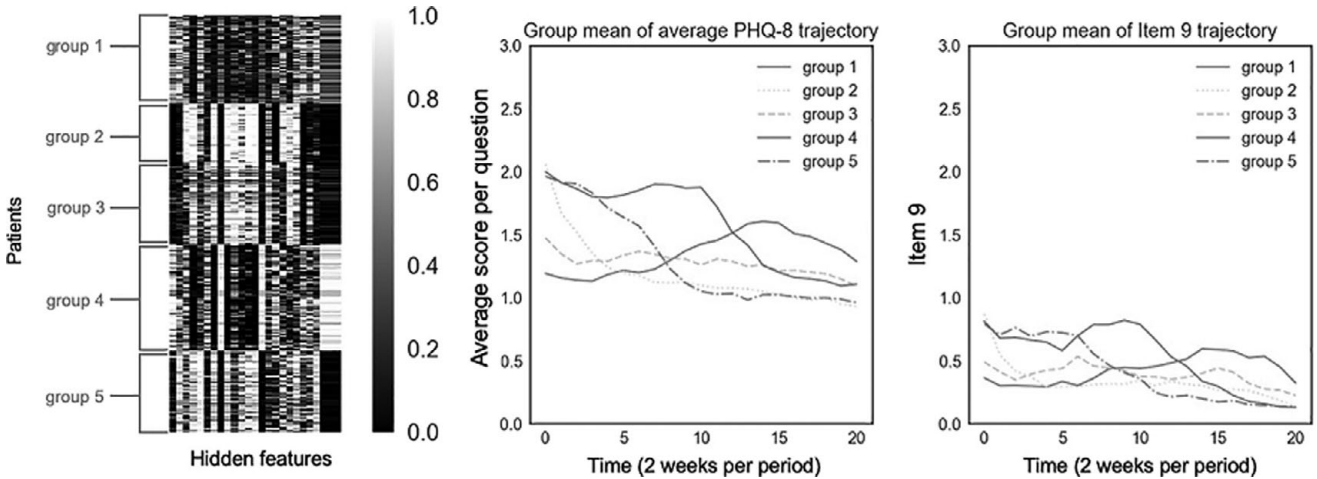


Figure 2.3 Clustering analysis for PHQ-8 (PHQ-9 total score minus the 9th question) found five subgroups in the learned features. Graphs show the mean scores of each subgroup for PHQ-8 and Item 9 (9th question on suicide ideation) scores over 20 biweeks [35].

Table 2.2. *Top rules identified by the Rulefit Model [39]*

	Decreasing risk rules		Increasing risk rules
Rule 1	Deepest increase between consecutive PHQ9 scores <7.50 & 75 percentile of PHQ9 score <14.62	Rule 7	Observing density >0.03 & Minimal PHQ9 score >8.50
Rule 2	25 percentile of PHQ9 score <6.13 & Volatility of PHQ9 score <9.64	Rule 8	Minimal PHQ9 score >9.50 & Volatility of difference between nearby PHQ9 scores <4.75
Rule 3	75 percentile of PHQ9 score <15.88 & Percentage of moderate depression <0.39	Rule 9	Latest PHQ9 score >17.50 & Volatility of PHQ9 score <7.33
Rule 4	Deepest decrease between consecutive PHQ9 scores >2.50 & 75 percentile of PHQ9 score <14.12	Rule 10	Minimal PHQ9 score >6.50 & 75 percentile of PHQ9 score > 14.88
Rule 5	Sex is male & Mean of 9th question scores <0.71 & Percentage of moderately severe <0.38	Rule 11	Age <65 & Percentage of severe depression >0.23
Rule 6	Latest PHQ9 score <8.50 & Maximal PHQ9 score <16.50	Rule 12	Deepest decrease between consecutive PHQ9 scores <13.50 & Mean of PHQ9 scores >14.73

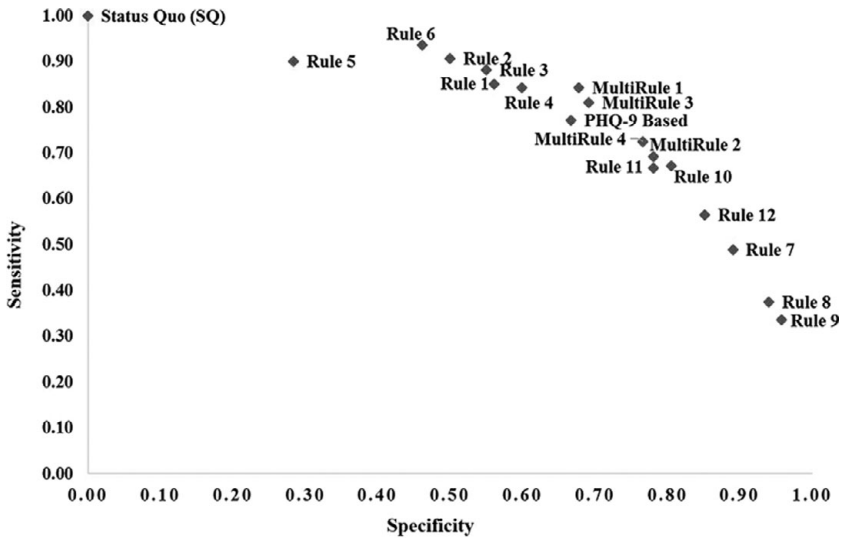


Figure 2.4 Comparison of monitoring accuracy for all strategies [39]

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

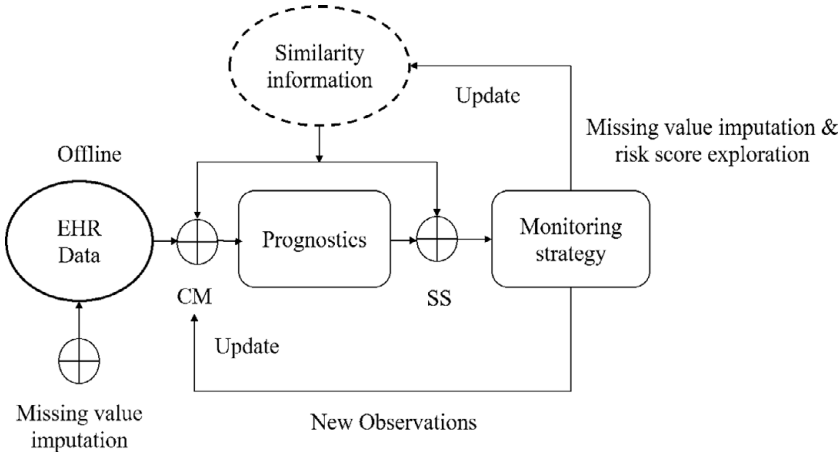


Figure 2.5 Overview: the CM-based prognosis and selective sensing for monitoring a heterogeneous patient population [38]

trajectories during online adaptive assessment and schedule the next visit under a capacity constraint. To do so, we integrated prognosis and sensing methodologies. For prognosis, a CM approach was used to predict the future progression risk of individuals. For adaptive monitoring, a selective sensing (SS) approach was developed to allocate limited sensing resources to monitor the high-risk individuals. See Figure 2.5.

We briefly describe the selective sensing method here. SS is formulated as an integer programming problem: the goal is to optimally allocate the sensing (monitoring) resources to detect a subgroup of high-risk individuals at each monitoring period. A patient's disease progression is modeled using the Markov model. Using a CM approach, we first estimated a transition matrix $\mathbf{P}_{it}$ to predict the progression risk of each individual patient i at each monitoring/sensing period t. Healthcare systems are often constrained by providers' capacity for mental health follow-up visits. For example, we learned that the demand for psychotherapy visit is much greater than supply at several health systems. We would prefer an algorithm that can optimally allocate the limited monitoring resources to detect high-risk individuals that are most likely to benefit from further interventions.

Suppose that there are N individuals and we denote the measurements of these individuals at each monitoring period as $\boldsymbol{x}_t = [x_{1t}, \ldots, x_{Nt}]$, where each measurement is a PHQ-9 score of the patient. We were interested in detecting the high-risk patients. Due to the limited sensing resources, we can only observe M out of N individuals at each period $(M < N)$. We introduced the

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

binary decision variable δ_{it} for each measurement x_{it} such that $\delta_{it} = 1$ if and only if x_{it} is observed at period t. Thus, the problem is how to choose δ_{it} at each period such that the sensing constraint is satisfied and highest-risk individuals are detected. Detailed technical formulation can be found in Lin et al. 2018 [38].

New observations collected by the sensing strategy at the next period were further incorporated in the CM-based prognostic method to update the prognosis of all individuals, guiding the monitoring decision in the next period, i.e., the CM uses these observations to update the canonical models and membership vector $\boldsymbol{c}_i$ for all individual models. Adaptively monitoring the predicted high-risk individuals may lead to an increased number of missing values on predicted low-risk individuals. To guarantee an accurate estimation of the next-period major depression risk, we imputed the missing value before running risk prediction.

We applied the CM and SS methods on an EHR dataset of 610 patients that have at least 6 observations within 40 weeks under ongoing treatment [38]. We characterized patients' depression progressions using Markov models and predicted the risk to severe depression using CM. We ran the selective sensing algorithm to adaptively monitor all patients over 15 time periods (each consists of two weeks) under a sensing capability of 100 patients in each period. Prediction performance is evaluated by the correlation between predicted risks of severe depression and ground truth risks (derived from observed depression onset of the patients). Detection performance is measured by the percentage of severe patients being detected and the average true risk among detected patients. For instance, when only 10% of the patients could be monitored each time, our results showed that the selective sensing algorithm outperformed the rank and selection method.

2.3.2.2 Adaptive Treatment Selection

Adaptive treatment design aims to optimally select a series of treatments to improve the health outcomes of depression patients. Adaptive treatments are personalized based on patient characteristics, behaviors, disease history, and response to treatment [40]. Decisions on treatment may include medications, drug doses, administering schedules, behavioral interventions, or no further treatment [40]. MDPs [41], reinforcement learning [42], and multi-armed bandits [43] are among the most widely used tools. Many challenges still remain, including insufficient knowledge of personal progression dynamics and learning individual response to treatments.

Study Five: Partially Observable CM and POMDP The aforementioned studies assumed fully observable disease state and no treatment feedback. In

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

Gong and Liu 2019 [44], we proposed a partially observable collaborative modeling (POCM) method. Depression is modeled using a Hidden Markov Model (HMM). In an HMM, disease progression is represented using transition probabilities between true disease states, and emission probabilities are probabilities of observing some measurements of the true disease states. Similar to the CM, each patient's transition matrix and emission matrix in the HMM are assumed to be a linear combination of several canonical progression groups in the population. The weight of each patient belonging to each group is called the membership. We developed a POCM solution algorithm to estimate the parameters of the transition and emission matrices.

Next, we used a partially observable MDP (POMDP) to make a sequence of adaptive treatment decisions based on the estimated dynamics. The hidden states of the POMDP are depression severities, observations are depression assessment scores (i.e., PHQ-9), and decisions are treatment options. The objective is to optimally select between two types of treatments in each time period and maximize health outcome over time. For example, Treatment I is usual care under antidepressant medications, and Treatment II is an intensive depression management program with telephone-based treatment coordination. The objective is to maximize discounted total rewards measured using Net Monetary Benefit (defined as total health benefits × willingness-to-pay – total cost). Health benefits can be measured by quality-adjusted life years gained.

The process of creating the adaptive treatment framework includes three steps. See Figure 2.6. (i) In the **learning step**, the canonical transition and

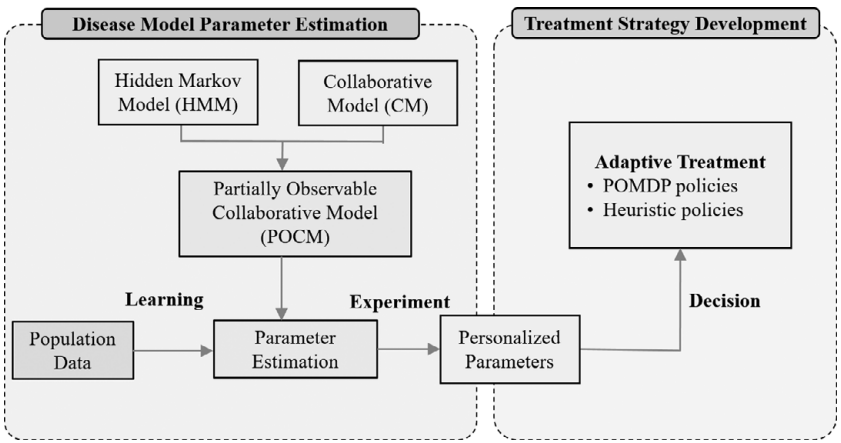


Figure 2.6 Overview of adaptive treatment design [44].

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

emission matrices for each progression pattern and patients' memberships are learned from an existing data set of patients under Treatment I using POCM. The population average treatment effect for Treatment II is assumed to be known from clinical trials, and such knowledge is used to estimate the canonical groups' parameters for Treatment II. (ii) In the **experiment step**, the personal dynamic for a new patient is initialized using model parameters estimated from the learning step and updated under both treatment options by running separate short trial periods; this is again accomplished by the POCM algorithm. The membership is then solved for each new patient belonging to each canonical group under either Treatment I or II. (iii) In the **decision step**, the optimal treatment strategy is obtained by solving a POMDP with the dynamic programming method (e.g., modified Monahan's algorithm [45]).

We compared the performance of the POMDP-based policies and several heuristic rule-based policies using a simulated depression treatment population. Results showed that the POCM can provide a better estimation of personal disease progression than the traditional method of solving an HMM when there are subgroup structures in the disease progression. We also demonstrated that the POCM-POMDP policies give the highest benefit for patients over the course of treatment. In addition, the POCM-based policies switch treatment less frequently than other policies. This research helps to advance the development of AI decision support tools for chronic disease care [44].

2.3.3 Cost-Effectiveness Analysis

A key missing link is how to estimate the long-term outcomes of these data-driven AI-assisted interventions in real-world clinical settings. Cost-effectiveness analyses are economic studies that use decision-analytic models to compare the costs and benefits of alternative interventions [46]. CEA models are useful for exploring alternative scenarios, extrapolating from intermediate endpoints to downstream outcomes, and informing decisions in the absence of data [47].

Study Six: Prognostic-Based CEA Prognostic-based monitoring that stratifies the individual's disease progression risk into different levels and adaptively allocates monitoring resource to high-risk individuals has the potential to improve patient health outcome and cost-effectiveness of the monitoring service. However, challenges include how to best apply prognostic models to inform the design of monitoring strategies and identify the cost-effective strategies.

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

In Lin et al. 2019 [48], we developed a decision support framework that integrated individual prognostics, monitoring strategy design and cost-effectiveness analysis (Figure 2.7). We applied the proposed framework to simulate the adaptive monitoring of a depression treatment population from EHR data. Several prediction algorithms with increasing complexity, including natural history matching, logistic regression, rule-based method, and Markov-based CM, were simulated to monitor the high-risk individuals for severe depression over time. We found six cost-effective monitoring strategies and demonstrated that the two routine monitoring strategies were dominated by the prognostic-based monitoring strategies (Figure 2.8). Methods from this

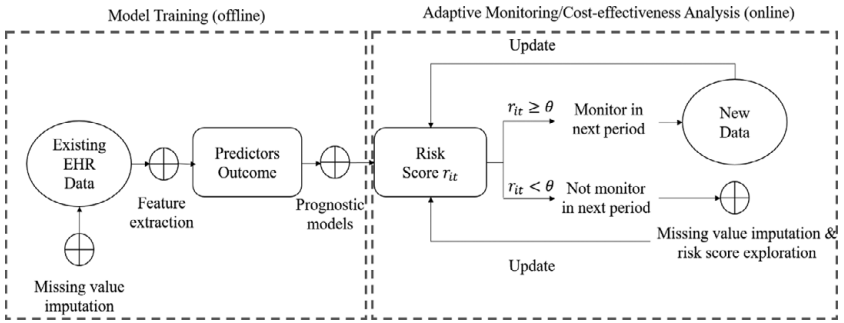


Figure 2.7 The framework of prognostic-based monitoring. Here, r_{it} denotes the risk score of individual i in tth monitoring period, and θ represents the threshold for monitoring [48].

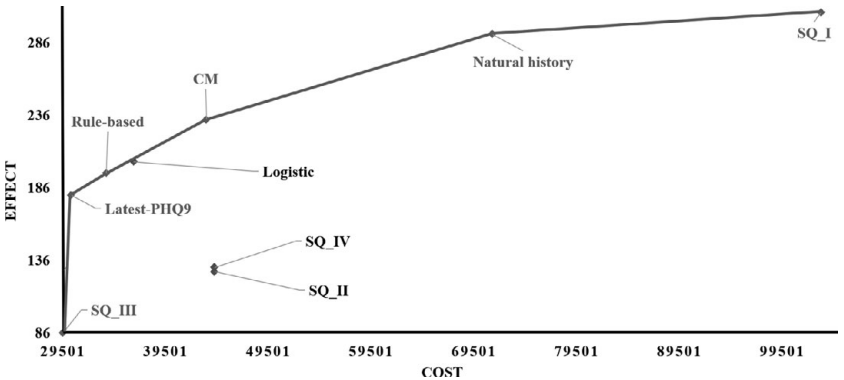


Figure 2.8 Cost-effectiveness frontier. The cost-effective strategies are represented by blue dots and the dominated strategies are denoted by red dots. SQ: status quo of routine monitoring under various frequencies [48].

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

research showed promise to implement prognostic-based monitoring of chronic conditions in clinical practice.

Study Seven: Cohort-Based CEA In Sun and Liu 2020 [49], we evaluated the cost-effectiveness of an adaptive remote monitoring technology for optimally switching between nine depression treatment lines. We used Markov-cohort models to simulate chronic depression patients' disease progression under monitoring and treatment over two years. Cohorts are defined by age (base case, 45 years), and sex (base case, 69% female). In addition, we considered heterogeneous disease progression patterns and clustered patients into three groups including a high-risk, a medium-risk, and a low-risk group of major depression. We considered five strategies: an adaptive technology that schedules follow-up appointment for treatment switch based on remotely monitoring patients' response to treatment with inaccuracy (i.e., imperfect sensitivity and specificity); a rule-based follow-up strategy that assigns the next follow-up time based on the patient's health state observed at the current follow-up (similar to current practice); and fixed frequency follow-up at every two-month, four-month, and six-month period. In the base case, since the monitoring accuracy and usage cost are uncertain, we simulated more than 1,000 scenarios and investigated how sensitivity, specificity, and cost of the remote monitoring technology would affect its cost-effectiveness.

Results showed for an adaptive remote monitoring technology with 0.75 sensitivity and specificity, it is cost effective with an incremental cost-effectiveness ratio ranging from $52,600/QALYs to $63,800/QALYs (2019 USD) gained compared to the rule-based follow-up strategy. Sensitivity analyses indicated that the imperfect technology is 63–78% cost effective depending on the risk group. In summary, a combination of methods including clustering, Markov-cohort model, and treatment simulation can be generalized to evaluate the cost-effectiveness of adaptive monitoring technology in other disease applications.

2.4 Summary

In this chapter, we presented several methodological challenges in advancing technology development for chronic depression and proposed three steps to achieve an AI-assisted technology framework. We showed highlights from seven research studies to aid in the design of adaptive monitoring and treatment strategies for chronic depression. This body of work include modeling complex disease trajectories, identifying patients at high risk of progression

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

through predictive analytics, and recommending adaptive and cost-effective follow-up care. These methods complement and build on each other to achieve the final goal of maximizing patient health outcomes. We believe AI-assisted technology in mental health holds the promise of transforming the current reactive practice to proactive and personalized monitoring and treatment, providing value to healthcare providers, patients, and healthcare systems. Given the surge in interest in telemedicine during and post the COVID19 pandemic, and the opportunity of telemedicine to treat depression, AI-assisted technology has an increasingly important role to play in patient care.

Methods presented in this chapter are advanced and not yet implemented in clinical practice. We hope that they may serve as fundamental building blocks for future AI-assisted technology to be translated into clinical practice. Though there is growing interest from academia, large healthcare systems, and private entrepreneurs to develop applications in remote monitoring and digital therapeutics, and some efforts have shown early success [50], the effectiveness and cost-effectiveness of these novel technologies still need to be proven and evaluated in long-term observational studies and comparative effectiveness trials. Ultimately, implementation success will depend on user acceptance, ease of use, system maintenance, and integration with EHR and the workflow of routine mental health services.

References

1. Centers for Disease Control and Prevention: Depression. www.cdc.gov/nchs/fastats/depression.htm.
2. Centers for Disease Control and Prevention: Depression morbidity. www.cdc.gov/nchs/data/nhis/earlyrelease/EarlyRelease202009-508.pdf.
3. National Institute of Mental Health: Depression. www.nimh.nih.gov/health/topics/depression/index.shtml. 2018
4. Centers for Disease Control and Prevention: Facts about suicide. www.cdc.gov/suicide/facts/index.html.
5. Insel, T.: Thomas Insel: toward a new understanding of mental illness. TED Talk. www.youtube.com/watch?v=PeZ-U0pj9LI. 2013
6. National Institute of Mental Health: Suicide. www.nimh.nih.gov/health/statistics/suicide.
7. Klein M. Bloomberg visual data: how Americans die. www.bloomberg.com/dataview/2014-04-17/how-americans-die.html. 2014.
8. Ribeiro J, Franklin J, Fox KR, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychological Medicine*. 2016; **46**(2):225–236.
9. USPSTF. Screening for depression in adults: U.S. preventive services task force recommendation statement. *Annals of Internal Medicine*. 2009; **151** (11):784–792.

https://doi.org/10.1017/9781108872188.005 Published online by Cambridge University Press

10. Parekh AK, Barton MB. The challenge of multiple comorbidity for the US health care system. *JAMA*. 2010; **303**(13):1303–1304.
11. Simon GE, VonKorff M, Rutter C, Wagner E. Randomised trial of monitoring, feedback, and management of care by telephone to improve treatment of depression in primary care. *BMJ*. 2000; **320**(7234):550–554.
12. Reynolds CFR, Frank E. US Preventive Services Task Force Recommendation Statement on Screening for Depression in Adults: not good enough. *JAMA Psychiatry*. 2016; **73**(3):189–190.
13. Pratt L, Brody D, Gu Q. *Antidepressant Use in Persons Aged 12 and Over: United States, 2005–2008*. NCHS Data Brief, No. 76. Hyattsville, MD: National Center for Health Statistics. 2011.
14. Shechter SM, Bailey MD, Schaefer AJ, Roberts MS. The optimal time to initiate HIV therapy under ordered health states. *Operations Research*. 2008; **56**(1):20–33.
15. Alagoz O, Maillart LM, Schaefer AJ, Roberts MS. The optimal timing of living-donor liver transplantation. *Management Science*. 2004; **50**(10):1420–1430.
16. Alagoz O, Maillart LM, Schaefer AJ, Roberts MS. Determining the acceptance of cadaveric livers using an implicit model of the waiting list. *Operations Research.* 2007; **55**(1):24–36.
17. Burhaneddin S, Maillart, LM, Shaefer, AJ, Alagoz O. Estimating the patient's price of privacy in liver transplantation. *Operations Research*. 2008; **56**(6):1393–1410.
18. Sutin AR, Terracciano A, Milaneschi Y, et al. The trajectory of depressive symptoms across the adult life span. *JAMA Psychiatry*. 2013; **70**(8):803–811.
19. Gunn J, Elliott P, Densley K, et al. A trajectory-based approach to understand the factors associated with persistent depressive symptoms in primary care. *Journal of Affective Disorders*. 2013; **148**(2-3):338–346.
20. Wang X, Sontag D, Wang F. Unsupervised learning of disease progression models. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2014:85–94.
21. Brandeau ML, Sainfort F, Pierskalla WP. *Operations Research and Health Care: A Handbook of Methods and Applications*. Boston: Kluwer Academic 2004.
22. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA*. 1996; **276** (15):1253–1258.
23. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices – modeling studies. *Value in Health*. 2003; **6** (1):9–17.
24. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. *Journal of Clinical Epidemiology,* 2012; **65**(10):1078–1087.
25. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics*, 2002; 11(1):33–42.
26. Musliner KL, Munk-Olsen T, Eaton WW, Zandi PP. Heterogeneity in long-term trajectories of depressive symptoms: patterns, predictors and outcomes. *Journal of Affective Disorders*, 2016; **192**:199–211.
27. Bishop CM. *Pattern Recognition and Machine Learning*. New York: Springer 2006.

28. Russell SJ, Norvig P, Davis E. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Prentice Hall 2010.
29. Drummond M. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford: Oxford University Press 2015.
30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*. 2001; **16**(9):606-613.
31. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases.* 1987; **40**(5):373–383.
32. Boor CD. *A Practical Guide to Splines*: New York: Springer 1978.
33. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PloS One*. 2013; **8**(6):e66341.
34. Shechter S. When to initiate, when to switch, and how to sequence HIV therapies: a Markov decision process approach. Doctoral thesis, University of Pittsburgh, Pittsburgh, PA; 2006.
35. Gong J, Simon GE, Liu S. Machine learning discovery of longitudinal patterns of depression and suicidal ideation. *PloS One.* 2019; **14**(9):e0222665.
36. Lin Y, Huang S, Simon GE, Liu S. Analysis of depression trajectory patterns using collaborative learning. *Mathematical Biosciences*. 2016; **282**:191–203.
37. Lin Y, Liu K, Byon E, et al. A collaborative learning framework for estimating many individualized regression models in a heterogeneous population. *IEEE Transactions on Reliability*. 2018; **67**(1):328–341.
38. Lin Y, Liu S, Huang S. Selective sensing of a heterogeneous population of units with dynamic health conditions. *IISE Transactions*. 2018; **50**(12):1076–1088.
39. Lin Y, Huang S, Simon GE, Liu S. Data-based decision rules to personalize depression follow-up. *Scientific Reports*. 2018; **8**(1):5064.
40. Kosorok MR, Moodie EE. *Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine*. ASA-SIAM Series on Statistics and Applied Mathematics. Philadelphia: Society for Industrial and Applied Mathematics 2016.
41. Schaefer AJ, Bailey MD, Shechter SM, Roberts MS. Modeling medical treatment using Markov decision processes. In: Brandeau ML, Sainfort F, Pierskalla WP, eds. *Operations Research and Health Care: A Handbook of Methods and Applications*. Boston: Springer 2004:593–612.
42. Vincent RD, Pineau J, Ybarra N, Naqa IE. Chapter 16: Practical reinforcement learning in dynamic treatment regimes. In: *Adaptive Treatment Strategies in Practice* 2016:263–296.
43. Negoescu D, Bimpikis K, Brandeau M, Iancu DA. Dynamic learning of patient response types: an application to treating chronic diseases. *Management Science*. 2017; **64**(8):3469–3970.
44. Gong J, Liu S. Partially observable collaborative model for optimizing personalized treatment selection. Paper presented at the 40th SMDM Annual Meeting; 2018; Montreal, Quebec, Canada.
45. Monahan GE. A survey of partially observable Markov decision processes: theory, models, and algorithms. *Management Science*. 1982; **28**(1):1–16.

46. Gold M. Panel on cost-effectiveness in health and medicine. *Medical Care.* 1996; **34**(12 Suppl):DS197–DS199.
47. Buxton MJ, Drummond MF, et al. Modelling in economic evaluation: an unavoidable fact of life. *Health Economics*. 1997; **6**(3):217–227.
48. Lin Y, Huang S, Simon GE, Liu S. Cost-effectiveness analysis of prognostic-based depression monitoring. *IIE Transactions on Healthcare Systems Engineering*. 2019; **9**(1):41–54.
49. Sun X, Liu S. A decision-analytic framework to evaluate the cost-effectiveness of adaptive monitoring technology: the case of chronic depression. Paper presented at the 42nd Annual SMDM Annual Meeting; 2020; virtual.
50. Wu S, Vidyanti I, Liu P, et al. Patient-centered technological assessment and monitoring of depression for low-income patients. *Journal of Ambulatory Care Management.* 2014; **37**(2):138–147.