Published in final edited form as:
Biometrics. 2014 March ; 70(1): 53–61. doi:10.1111/biom.12132.

Set-valued dynamic treatment regimes for competing outcomes

Eric B. Laber[1,*], **Daniel J. Lizotte**[2,**], and **Bradley Ferguson**[1,***]
[1]Department of Statistics NC State University, Raleigh, NC 27695, USA
[2]Department of Computer Science, University of Waterloo, Ontario, N2L 3G1

Summary

Dynamic treatment regimes operationalize the clinical decision process as a sequence of functions, one for each clinical decision, where each function maps up-to-date patient information to a single recommended treatment. Current methods for estimating optimal dynamic treatment regimes, for example *Q*-learning, require the specification of a single outcome by which the 'goodness' of competing dynamic treatment regimes is measured. However, this is an over-simplification of the goal of clinical decision making, which aims to balance several potentially competing outcomes, e.g., symptom relief and side-effect burden. When there are competing outcomes and patients do not know or cannot communicate their preferences, formation of a single composite outcome that correctly balances the competing outcomes is not possible. This problem also occurs when patient preferences evolve over time. We propose a method for constructing dynamic treatment regimes that accommodates competing outcomes by recommending sets of treatments at each decision point. Formally, we construct a sequence of set-valued functions that take as input up-to-date patient information and give as output a recommended subset of the possible treatments. For a given patient history, the recommended set of treatments contains all treatments that produce non-inferior outcome vectors. Constructing these set-valued functions requires solving a non-trivial enumeration problem. We offer an exact enumeration algorithm by recasting the problem as a linear mixed integer program. The proposed methods are illustrated using data from the CATIE schizophrenia study.

Keywords

Dynamic Treatment Regimes; Personalized Medicine; Composite Outcomes; Competing Outcomes; Preference Elicitation

1. Introduction

Dynamic treatment regimes (DTRs) operationalize the clinical decision-making process wherein a clinician selects a treatment based on current patient characteristics and then continues to adjust treatment over time in response to the evolving health status of the patient. A DTR is a sequence of decision rules, one for each decision point. Each rule takes as input current patient information and gives as output a recommended treatment. There is growing interest in estimating "optimal" DTRs from randomized or observational data. A

*laber@stat.ncsu.edu
**dlizotte@uwaterloo.ca
***bradtferguson@gmail.com

Supplementary Materials
Web Appendix A, referenced in Sections 1, 3, and 5, Web Appendix B, referenced in Section 3, Web Appendix C, referenced in Section 4, Web Appendix D, referenced in Section 4, and computer code implementing the SVDTR algorithm described in Section 3.1, are available with this paper at the *Biometrics* website on Wiley Online Library.

DTR is said to be optimal if, when applied in the population of interest, it maximizes the average clinical outcome. Optimal DTRs have been estimated for chronic conditions including ADHD (Laber et al., 2011; Nahum-Shani et al., 2010, 2012), depression (Schulte et al., 2012; Song et al., 2012), HIV infection (Moodie et al., 2007), schizophrenia (Shortreed et al., 2011), and cigarette addiction (Strecher et al., 2006). Approaches for estimating optimal DTRs from data include *Q*-learning (Watkins and Dayan, 1992; Nahum-Shani et al., 2010), *A*-learning (Murphy, 2003; Blatt et al., 2004; Robins, 2004), regret regression (Henderson et al., 2010), and direct value maximization (Orellana et al., 2010; Zhang et al., 2012; Zhao et al., 2012).

To estimate a DTR from data using any of the above methods, one must specify a single outcome and neglect all others. For example, one might seek the most effective DTR without regard for side-effects. Alternatively, one could form a linear combination of two outcomes, e.g., side effects and effectiveness, yielding a single composite outcome. Forming this outcome requires the elicitation of a trade-off between two outcomes; for example, one would need to know that a gain of 1 unit of effectiveness is worth a cost of 3 units of side-effects. However, for some illnesses, e.g., severe schizophrenia, preferences across outcomes can vary widely across patients (Kinter, 2009). Thus, even if one could elicit this trade-off at an aggregate level, assuming that a particular trade-off holds for all decision-makers is not reasonable since each will have his or her own individual preferences which cannot be known *a priori*. Furthermore, patients may not know their preferences, they may be unable to communicate them, or they may have preferences which evolve over time (Strauss et al., 2011).

Lizotte et al. (2012) present one approach to dealing with this problem using a method that estimates an optimal DTR for all possible linear trade-offs simultaneously. Their method can also be used to explore what range of trade-offs is consistent with each available treatment. Nonetheless, their method assumes that any outcome preference can be expressed by a composite outcome that is a *linear* combination of the outcomes under consideration. They still (perhaps implicitly) require the decision-maker to assess and reason about the space of linear composite outcomes. In addition, their approach suggests actions based on the assumption that preferences remain fixed over time.

We propose *set-valued Dynamic Treatment Regimes* (SVDTRs) as an alternative to DTRs that accommodates competing outcomes and preference heterogeneity both across patients and time, but avoids eliciting trade-offs between outcomes. Like a DTR, an SVDTR is a sequence of decision rules. However, the decision rules that compose an SVDTR take as input current patient information and give as output a *set* of recommended treatments. This set is a singleton when there exists a treatment that is best across all outcomes but contains multiple treatments otherwise. Treatments that are inferior according to all outcomes are eliminated. By presenting multiple reasonable treatments, our proposed method still allows for the incorporation of clinical judgment, individual patient preferences (to the extent that they are known), cost, and local availability, when deciding among the non-inferior treatments. Our approach does not *require* any individual preference information from the decision maker; however, in its most general form, our approach makes use of an elicited ‘clinically significant’ difference on each outcome scale to help decide if one treatment is clearly inferior to another (see Friedman et al., 2010, for example).

This work is motivated by the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study (Stroup et al., 2003), in which schizophrenic patients were randomized up to two times to different treatments. CATIE has three features that make it amenable to our proposed approach: i) It contains data we can use to individualize treatment. ii) It follows patients over multiple treatment phases. iii) It contains data on important competing

outcomes. The CATIE data include both measures of symptoms and side-effects, and it is well-established that treatments that provide some of the best symptom relief have the worst side-effects (Breier et al., 2005; Allison et al., 1999). Thus, to illustrate our approach we present an SVDTR-based analysis of CATIE in Section 4.

Our primary contribution is the introduction of SVDTRs, which offer a new approach to operationalizing sequential clinical decision making that is informed by predicted competing outcomes and by clinical judgment. We also provide a novel mathematical programming formulation which gives a computationally efficient method to estimate SVDTRs from data. In Section 2, we review the *Q*-learning algorithm for estimating optimal DTRs from data. In Section 3 we propose an SVDTR for the two decision point problem, and in Section 3.1 we describe our mathematical programming approach for estimating an SVDTR from data. Section 4 presents our analysis of CATIE. For clarity, the main body of the paper considers only binary treatment decisions; we give an extension to an arbitrary number of treatments in Web Appendix A.

2. Single outcome decision rules

In this section we review the *Q*-learning algorithm for estimating an optimal DTR when there is a single outcome of interest. For simplicity, we consider the case in which there are two decision points and two treatment options at each decision point. In this setting the data available to estimate an optimal DTR consists of n trajectories $(\mathbf{H}_1, A_1, \mathbf{H}_2, A_2, Y)$, one for each patient, drawn *i.i.d.* from some unknown distribution. We use capital letters like $\mathbf{H}_1$ and A_1 to denote random variables and lower case letters like $\mathbf{h}_1$ and a_1 to denote realized values of these random variables. The components of each trajectory are as follows: $\mathbf{H}_t \in \mathbb{R}^{p_t}$ denotes patient information collected prior to the assignment of the tth treatment, and thus is information the decision maker can use to inform the tth treatment decision (note that $\mathbf{H}_2$ may contain some or all of the vector $(A_1, \mathbf{H}_1^\top)^\top$); $A_t \in \{-1, 1\}$ denotes the tth treatment assignment; $Y \in \mathbb{R}$ denotes the outcome of interest which is assumed to be coded so that higher values are more desirable than lower values. The outcome Y is commonly a measure of treatment effectiveness, but could also be a composite measure attempting to balance different objectives. Given the definition of Y, the goal is to construct a pair of decision rules $\pi = (\pi_1, \pi_2)$ where $\pi_t(\mathbf{h}_t)$ denotes a decision rule for assigning treatment at time t to a patient with history $\mathbf{h}_t$ in such a way that the expected response Y, given such treatment assignments, is maximized. Formally, if E^π denotes the joint expectation over $\mathbf{H}_t$, A_t, and Y under the restriction that $A_t = \pi_t(\mathbf{H}_t)$, then the optimal decision rule π^{opt} satisfies $\mathrm{E}^{\pi^{\mathrm{opt}}} Y = \sup_\pi \mathrm{E}^\pi Y$. Note this definition of optimality ignores the impact of the DTR π^{opt} on any outcome not incorporated into Y.

One method for estimating an optimal DTR is the *Q*-learning algorithm (Watkins and Dayan, 1992). *Q*-learning is an approximate dynamic programming procedure that relies on regression models to approximate the conditional expectations $Q_2(\mathbf{h}_2, a_2) \triangleq \mathrm{E}(Y|\mathbf{H}_2=\mathbf{h}_2, A_2=a_2)$, and $Q_1(\mathbf{h}_1, a_1) \triangleq \mathrm{E}(\max_{a_2\in\{-1,1\}} Q_2(\mathbf{H}_2, a_2)|\mathbf{H}_1=\mathbf{h}_1, A_1=a_1)$. The function Q_t is termed the *stage-t Q-function*. The function $Q_2(\mathbf{h}_2, a_2)$ measures the quality of assigning treatment a_2 at the second decision point to a patient with history $\mathbf{h}_2$. The function $Q_1(\mathbf{h}_1, a_1)$ measures the quality of assigning treatment a_1 at the first decision point to a patient with history $\mathbf{h}_1$, assuming optimal treatment decisions will be made at the second decision point. Hence, it follows that $\pi_t^{\mathrm{opt}}(\mathbf{h}_t) = \arg\max_{a_t\in\{-1,1\}} Q_t(\mathbf{h}_t, a_t)$, $t = 1, 2$. This is the *dynamic programming* solution to finding the optimal sequence of decision rules (Bellman, 1957).

In practice, the *Q*-functions are not known and must be estimated from data. We consider linear working models of the form $Q_t(\mathbf{h}_t, a_t) = \mathbf{h}_{t,1}^\top \beta_t + a_t \mathbf{h}_{t,2}^\top \psi_t$, where $\mathbf{h}_{t,1}$ and $\mathbf{h}_{t,2}$ are (possibly the same) vector summaries of $\mathbf{h}_t$. Note that $\mathbf{h}_{t,j}$, $j = 1, 2$ might contain polynomial terms or other basis expansions as appropriate. The *Q*-learning algorithm proceeds in three steps:

1. Estimate the parameters of the working model for the stage-2 *Q*-function using least squares. Let $\hat{\beta}_2$ and $\hat{\psi}_2$ denote these estimates, and let $\hat{Q}_2(\mathbf{h}_2, a_2)$ denote the fitted model.
2.
   a. Define the predicted future outcome $\tilde{Y}$ following the estimated optimal decision rule at stage two as $\tilde{Y} \triangleq \max_{a_2 \in \{-1,1\}} \hat{Q}_2(\mathbf{H}_2, a_2)$.
   b. Estimate the parameters indexing the working model for the stage-1 *Q*-function using least squares. That is, regress $\tilde{Y}$ on $\mathbf{H}_1$ and A_1 using the working model to obtain $\hat{\beta}_1$ and $\hat{\psi}_1$. Let $\hat{Q}_1(\mathbf{h}_1, a_1)$ denote the fitted model.
3. The *Q*-learning estimate of π^{opt} is $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2)$ where $\hat{\pi}_t(\mathbf{h}_t) = \arg\max_{a_t \in \{-1,1\}} \hat{Q}(\mathbf{h}_t, a_t)$.

The *Q*-learning algorithm is simple to implement and easy to interpret given its connections to regression. Therefore, we use *Q*-learning as the basis for developing SVDTRs. Alternatives to *Q*-learning are listed in Section 1.

3. Set-valued dynamic treatment regimes

In a DTR, the optimal decision rule at time $t = 1$ depends critically on the decision rule that will be used at time $t = 2$, which in turn depends on the *Q*-functions at time $t = 2$. This is why *Q*-learning and related methods use backwards recursive estimation beginning at the final stage. Thus, if we cannot estimate the *Q*-function at $t = 2$ for any reason, existing recursive approaches like *Q*-learning cannot be applied. It follows that if the optimal rule for $t = 2$ depends on preference, but preference information is unavailable, *Q*-learning cannot be directly applied and we must devise a new strategy.

In some populations, e.g., severe schizophrenics, high quality preference elicitation may not be possible (Kinter, 2009; Strauss et al., 2011), which can lead to misspecification of the composite outcome needed to estimate *Q*-functions (see Web Appendix B for an illustration of how composite outcome misspecification can impact the quality of a decision rule). Furthermore, preference may evolve unpredictably over time so even if patient preferences were known exactly at the time of each treatment decision, *future* treatment preferences are unknown, and this precludes backwards recursive estimation.

Table 1 illustrates the foregoing problems in simplified setting with two hypothetical subjects drawn from different populations, no subject covariates, and two competing outcomes generically termed ‘side-effects’ and ‘efficacy.’ For Subject A, an initial preference for efficacy suggests treatment 1 at the first stage. Suppose, however, that during the course of the first treatment Subject A develops a strong aversion to side-effects. Because the initial treatment was chosen assuming a static preference for efficacy, Subject A is left with poor and very poor choices in terms of their current preference at the second stage. Given the information provided in the table the decision maker may recommend treatment -1 initially to allow for better second stage treatment choices; however, it is important to note that once estimates of outcomes and viable treatment strategies are provided to the decision maker (see below) treatment choice is no longer a statistical problem. An alternative strategy would be to apply the *Q*-learning algorithm with respect to each outcome. In the case of Subject A, the *Q*-learning algorithm for side-effects would

recommend treatment -1, whereas the *Q*-learning algorithm for efficacy would recommend treatment 1. The disagreement between the output of these algorithms could serve as signal to the decision maker that other external factors (e.g., clinical judgment, past treatment preferences, etc.) should be incorporated as 'tie-breakers.' However, Subject B in Table 1 demonstrates that this strategy will not work in general. The *Q*-learning algorithm for side-effects and efficacy both recommend treatment 1 at the first stage for Subject B. However, treatment 1 at the first stage leads to extreme and potentially undesirable trade-offs at the second stage.

In this section we propose set-valued DTRs for two decision points and two competing outcomes. An extension to an arbitrary number of treatments is given in Web Appendix A. The data available to estimate a pair of decision rules, one for each patient, comprises n trajectories ($\mathbf{H}_1, A_1, \mathbf{H}_2, A_2, Y, Z$) drawn *i.i.d.* from a fixed but unknown distribution. The first four elements in each trajectory are the same as the *Q*-learning setup and $Z, Y \in \mathbb{R}$ denote competing outcomes observed sometime after the assignment of the second treatment A_2. We assume that both Y and Z are coded so that higher values are preferred.

Our method can make use of *clinically significant differences* $\Delta_Y \geq 0$ and $\Delta_Z \geq 0$ for outcomes Y and Z respectively, to differentiate between treatment outcomes. We call a difference in outcome *clinically significant* if a clinician would be willing to change her treatment strategy given that this change was expected to yield a difference of at least Δ_Y (Δ_Z) in the outcome Y (Z), all else being equal. These differences may be elicited from a panel of experts, estimated from historical data, or taken from existing clinical guidelines. Importantly, in eliciting Δ_Y there is no need to reference the competing outcome Z, and vice versa when eliciting Δ_Z. They may be patient-independent and constant over time. We believe the incorporation of clinically significant differences adds to the utility and interpretability of our approach in many domains; nevertheless, they are not necessary for the validity of our algorithms and could be taken to both equal zero. Furthermore, our algorithms do not preclude clinical significances being functions of individual patient characteristics, being different at each time point, or allowing dependence between Δ_Y and Δ_Z, however, we do not incorporate these generalizations into our notation. To avoid having to repeatedly qualify our discussion, we will assume that both $\Delta_Y > 0$ and $\Delta_Z > 0$.

The goal is to construct a pair of decision rules $\pi = (\pi_1, \pi_2)$ where $\pi_t : \mathbb{R}^{p_t} \to \{\{-1, 1\}, \{-1\}, \{1\}\}$ maps up-to-date patient information to a subset of the possible decisions. Ideally, for a patient presenting with $\mathbf{h}_2$ at the second stage the set-valued decision rule would recommend a *single* treatment if that treatment is expected to yield a clinically significant improvement (relative to the alternative treatment) in at least one of the outcomes and, in addition, that treatment is not expected to lead to a significant detriment in the other outcome. If the preceding condition does not hold for one of the treatments then the decision rule should return the set $\{-1, 1\}$ and leave the 'tie-breaking' to the decision maker. Define the (non-normalized) second stage treatment effects as
$r_{2Y}(\mathbf{h}_2) \triangleq \mathrm{E}(Y|\mathbf{H}_2=\mathbf{h}_2, A_2=1) - \mathrm{E}(Y|\mathbf{H}_2=\mathbf{h}_2, A_2=-1)$, and likewise
$r_{2Z}(\mathbf{h}_2) \triangleq \mathrm{E}(Z|\mathbf{H}_2=\mathbf{h}_2, A_2=1) - \mathrm{E}(Z|\mathbf{H}_2=\mathbf{h}_2, A_2=-1)$. Then, the ideal second stage decision rule, say $\pi_{2\Delta}^{\mathrm{Ideal}}$, is given by

$$\pi_{2\Delta}^{\mathrm{Ideal}}(\mathbf{h}_2) = \begin{cases} \{\mathrm{sgn}(r_{2Y}(\mathbf{h}_2))\}, & \text{if} |r_{2Y}(\mathbf{h}_2)| \geq \Delta_Y \text{ and } \mathrm{sgn}(r_{2Y}(\mathbf{h}_2)) r_{2Z}(\mathbf{h}_2) > -\Delta_Z, \\ \{\mathrm{sgn}(r_{2Z}(\mathbf{h}_2))\}, & \text{if} |r_{2Z}(\mathbf{h}_2)| \geq \Delta_Z \text{ and } \mathrm{sgn}(r_{2Z}(\mathbf{h}_2)) r_{2Y}(\mathbf{h}_2) > -\Delta_Y, \\ \{-1, 1\}, & \text{otherwise}, \end{cases} \quad (1)$$

where sgn denotes the signum function. Figure 1 illustrates how $\pi_{2\Delta}^{\mathrm{Ideal}}(\mathbf{h}_2)$ depends on $r_{2Y}(\mathbf{h}_2)$ and $r_{2Z}(\mathbf{h}_2)$, Δ_Y, and Δ_Z. If we consider the $(r_{2Y}(\mathbf{h}_2), r_{2Z}(\mathbf{h}_2)) \in \mathbb{R}^2$, its location

relative to the points (Δ_Y, Δ_Z), ($-\Delta_Y$, Δ_Z), (Δ_Y, $-\Delta_Z$) and ($-\Delta_Y$, $-\Delta_Z$) determines whether we prefer treatment 1, prefer treatment −1, or are undecided according to the foregoing criteria.

We now define $\pi_{1\Delta}^{\text{Ideal}}$ given that a clinician always selects treatments from the set-valued decision rule $\pi_{2\Delta}^{\text{Ideal}}$ at the second stage. This problem is complicated by the fact that, unlike in the standard setting, there exists a set of histories $\mathbf{h}_2$ at the second stage—those for which $\pi_{2\Delta}^{\text{Ideal}}(\mathbf{h}_2)=\{-1,1\}$—where we do not know which treatment would be chosen. To address this, we begin by assuming that some *non*-set-valued decision rule τ_2 will be used at the second stage, we will then consider an appropriate set of possible τ_2 in order to define $\pi_{1\Delta}^{\text{Ideal}}$.

Suppose a non-set-valued decision rule $\tau_2 : \mathbb{R}^{p2} \to \{-1, 1\}$ is used to assign treatments at the second stage. That is, a patient presenting with history $\mathbf{h}_2$ would be assigned treatment $\tau_2(\mathbf{h}_2)$. Define $Q_{2Y}(\mathbf{h}_2, \tau_2) \triangleq \mathrm{E}(Y|\mathbf{H}_2=\mathbf{h}_2, A_2=\tau_2(\mathbf{h}_2))$. Furthermore, define $Q_{1Y}(\mathbf{h}_1, a_1, \tau_2) \triangleq \mathrm{E}(Q_{2Y}(\mathbf{H}_2, \tau_2)|\mathbf{H}_1=\mathbf{h}_1, A_1=a_1)$ so that $Q_{1Y}(\mathbf{h}_1, a_1, \tau_2)$ is the expected outcome for a patient with first stage history $\mathbf{H}_1 = \mathbf{h}_1$ treated at the first stage with $A_1 = a_1$ and the decision rule τ_2 at the second stage. Replacing Y with Z yields $Q_{2Z}(\mathbf{h}_2, \tau_2)$ and $Q_{1Z}(\mathbf{h}_1, a_1, \tau_2)$. Thus, if it is known that a clinician will follow τ_2 at the second decision point, then the ideal decision rule at the first decision point is given by

$$\pi_{1\Delta}^{\text{Ideal}}(\mathbf{h}_1,\tau_2)=\begin{cases}\{\text{sgn}(r_{1Y}(\mathbf{h}_1,\tau_2))\}, & \text{if } |r_{1Y}(\mathbf{h}_1,\tau_2)| \geq \Delta_Y \\ & \text{and } \text{sgn}(r_{1Y}(\mathbf{h}_1,\tau_2))r_{1Z}(\mathbf{h}_1,\tau_2) > -\Delta_Z, \\ \{\text{sgn}(r_{1Z}(\mathbf{h}_1,\tau_2))\}, & \text{if} |r_{1Z}(\mathbf{h}_1,\tau_2)| \geq \Delta_Z \\ & \text{and } \text{sgn}(r_{1Z}(\mathbf{h}_1,\tau_2))r_{1Y}(\mathbf{h}_1,\tau_2) > -\Delta_Y, \\ \{-1,1\}, \text{otherwise}, & \end{cases} \quad (2)$$

where $r_{1Y}(\mathbf{h}_1,\tau_2) \triangleq Q_{1Y}(\mathbf{h}_1,1,\tau_2) - Q_{1Y}(\mathbf{h}_1,-1,\tau_2)$, and similarly $r_{1Z}(\mathbf{h}_1,\tau_2) \triangleq Q_{1Z}(\mathbf{h}_1,1,\tau_2) - Q_{1Z}(\mathbf{h}_1,-1,\tau_2)$. Note that $\pi_{1\Delta}^{\text{Ideal}}(\mathbf{h}_2,\tau_2)$ assigns a single treatment if that treatment is expected to yield a clinically significant improvement on one or both the outcomes while not causing clinically significant loss in either outcome *assuming the clinician will follow* τ_2 *at the second decision point.*

We now describe how to construct the ideal decision rule at the first decision point when the rule at the second decision point is set-valued. We say a *non*-set-valued rule τ_2 is *compatible* with a set-valued decision rule τ_2 if and only if

$$\tau_2(\mathbf{h}_2) \in \pi_2(\mathbf{h}_2) \ \ \forall \mathbf{h}_2 \in \mathbb{R}^{p_2}. \quad (3)$$

Let $C(\pi_{2\Delta}^{\text{Ideal}})$ be the set of all rules that are compatible with $\pi_{2\Delta}^{\text{Ideal}}$. We define $\pi_{1\Delta}^{\text{Ideal}}$ to be the set-valued decision rule

$$\pi_{1\Delta}^{\text{Ideal}}(\mathbf{h}_1) = \bigcup_{\tau_2 \in C(\pi_{2\Delta}^{\text{Ideal}})} \pi_{1\Delta}^{\text{Ideal}}(\mathbf{h}_1,\tau_2). \quad (4)$$

Our motivation for this definition is a desire to maintain as much choice as possible at stage 1, while making as few assumptions about future behaviour as possible. The definition in (4) assumes only that in the future some τ_2 in accordance with π_2 will be followed. Therefore at stage 1 we would only eliminate treatments for which there exists *no compatible future decision rule* that makes that treatment a desirable choice.

However, if we do not impose some smoothness constraints on τ_2, the set $C(\pi_{2\Delta}^{\text{Ideal}})$ can be very large, and computing the union (4) can become intractable. Furthermore, $C(\pi_{2\Delta}^{\text{Ideal}})$ may contain unreasonable future policies. Suppose that $\pi_{2\Delta}^{\text{Ideal}}(\mathbf{h}_2)=\{-1,1\}$ for all $\mathbf{h}_2$ in some non-null set H_2. Then the policy that assigns 1 to rational-valued histories in H_2 and −1 to irrational-valued histories in H_2 belongs to $C(\pi_{2\Delta}^{\text{Ideal}})$ even though it is clearly not a reasonable policy to follow. We will see that the modelling choices made to estimate Q_{2Y} and Q_{2Z} suggest a sensible subset of $C(\pi_{2\Delta}^{\text{Ideal}})$ over which to take the union (4) instead. We provide a mathematical programming formulation that allows us to use existing optimization algorithms to efficiently compute the union over this much smaller subset (see below).

We now turn to the estimation of $\pi_{1\Delta}^{\text{Ideal}}$ and $\pi_{2\Delta}^{\text{Ideal}}$ from data. As in the *Q*-learning setup, let $\mathbf{h}_{t,j}$ j = 1, 2 denote vector summaries of the history at time *t*. To estimate the ideal second stage decision rule we postulate linear models for second stage *Q*-functions, say, of the form $Q_{2Y}(\mathbf{h}_2,a_2)=\mathbf{h}_{2,1}^\top\beta_{2Y}+a_2\mathbf{h}_{2,2}^\top\psi_{2Y}$, $Q_{2Z}(\mathbf{h}_2,a_2)=\mathbf{h}_{2,1}^\top\beta_{2Z}+a_2\mathbf{h}_{2,2}^\top\psi_{2Z}$, which we estimate using least squares. In a slight abuse of notation, we write $\mathbf{h}_{t,j,i}$ to denote the *j*th vector summary (j = 1, 2) of history $\mathbf{h}_t$ (t = 1, 2) for subject i (i = 1, …, n), and $\mathbf{h}_{t,\cdot,i}$ to denote the history at time *t* for subject *i*. The estimated ideal second stage set-valued decision rule $\hat{\pi}_{2\Delta}$ is the plug-in estimate of (1). In order to estimate the ideal decision rule at the first decision point we must characterize how a clinician might assign treatments at the second decision point. We begin by assuming that clinicians' behavior, denoted by τ_2, is *compatible* with $\hat{\pi}_{2\Delta}$ as defined in (3), and we further assume that τ_2 can be expressed as a thresholded linear function of $\mathbf{h}_2$. We call such decision rules *feasible* for $\hat{\pi}_{2\Delta}$, and we define the set of feasible decision rules at stage 2 by

$F(\hat{\pi}_{2\Delta}) \triangleq \{\tau_2:\exists_\rho \in \mathbb{R}^{\rho_{2,2}} \text{ s.t. } \tau_2(\mathbf{h}_2)=\text{sgn}(\mathbf{h}_{2,2}^\top\rho) \text{ and } \tau_2 \in C(\hat{\pi}_{2\Delta})\}$. Here, $p_{2,2}$ = dim($\mathbf{h}_{2,2}$). This is exactly the set of all stage 2 decision rules that would be output by *Q*-learning for *some* outcome on the given space of histories.

Thus, $F(\hat{\pi}_{2\Delta})$ denotes the set of second stage non-set-valued decision rules that a clinician might follow if they were presented with $\hat{\pi}_{2\Delta}$. This set is indexed by the vector $\rho \in \mathbb{R}^{p2,2}$. It can be verified that $F(\hat{\pi}_{2\Delta})$ is non-empty since $\text{sgn}(\mathbf{h}_{2,2}^\top(\frac{1}{2\Delta_Y}\hat{\psi}_{2Y}+\frac{1}{2\Delta_Z}\hat{\psi}_{2Z}))$ belongs to $F(\hat{\pi}_{2\Delta})$. For an arbitrary $\tau_2 \in F(\hat{\pi}_{2\Delta})$, define the working models

$$\begin{aligned} Q_{1Y}(\mathbf{h}_1,a_1,\tau_2) &= \mathbf{h}_{1,1}^\top\beta_{1Y}(\tau_2)+a_1\mathbf{h}_{1,2}^\top\psi_{1Y}(\tau_2), \\ Q_{1Z}(\mathbf{h}_1,a_1,\tau_2) &= \mathbf{h}_{1,1}^\top\beta_{1Z}(\tau_2)+a_1\mathbf{h}_{1,2}^\top\psi_{1Z}(\tau_2), \end{aligned} \quad (5)$$

where $\beta_{1Y}(\tau_2)$, $\psi_{1Y}(\tau_2)$, $\beta_{1Z}(\tau_2)$, and $\psi_{1Z}(\tau_2)$ are coefficient vectors specific to τ_2. For a fixed τ_2 one can estimate these coefficients by regressing $\hat{Q}_{2Y}(\mathbf{H}_2,\tau_2)=\mathbf{H}_{2,1}^\top\hat{\beta}_{2Y}+\tau_2(\mathbf{H}_2)\mathbf{H}_{2,2}^\top\hat{\psi}_{2Y}$ and $\hat{Q}_{2Z}(\mathbf{H}_2,\tau_2)=\mathbf{H}_{2,1}^\top\hat{\beta}_{2Z}+\tau_2(\mathbf{H}_2)\mathbf{H}_{2,2}^\top\hat{\psi}_{2Z}$ on $\mathbf{H}_1$ and A_1 using the working models in (5). Let $\hat{Q}_{1Y}(\mathbf{h}_1, a_1, \tau_2)$ and $\hat{Q}_{1Z}(\mathbf{h}_1, a_1, \tau_2)$ denote these fitted models, and let $\hat{r}_{1Y} \triangleq \hat{Q}_{1Y}(\mathbf{h}_1,1,\tau_2)-\hat{Q}_{1Y}(\mathbf{h}_1,-1,\tau_2)$, and $\hat{r}_{1Z} \triangleq \hat{Q}_{1Z}(\mathbf{h}_1,1,\tau_2)-\hat{Q}_{1Z}(\mathbf{h}_1,-1,\tau_2)$. Define

$$\hat{\pi}_{1\Delta}(\mathbf{h}_1,\tau_2)=\begin{cases} \{\text{sgn}(\hat{r}_{1Y}(\mathbf{h}_1,\tau_2))\}, & \text{if } |\hat{r}_{1Y}(\mathbf{h}_1,\tau_2)| \geq \Delta_Y \text{ and } \text{sgn}(\hat{r}_{1Y}(\mathbf{h}_1,\tau_2))\hat{r}_{1Z}(\mathbf{h}_1,\tau_2) > -\Delta_Z, \\ \{\text{sgn}(\hat{r}_{1Z}(\mathbf{h}_1,\tau_2))\}, & \text{if } |\hat{r}_{1Z}(\mathbf{h}_1,\tau_2)| \geq \Delta_Z \text{ and } \text{sgn}(\hat{r}_{1Z}(\mathbf{h}_1,\tau_2))\hat{r}_{1Y}(\mathbf{h}_1,\tau_2) > -\Delta_Y, \\ \{-1,1\}, & \text{otherwise}, \end{cases} \quad (6)$$

and

$$\widehat{\pi}_{1\Delta}(\mathbf{h}_1) = \bigcup_{\tau_2 \in F(\widehat{\pi}_{2\Delta})} \widehat{\pi}_{1\Delta}(\mathbf{h}_1, \tau_2). \quad (7)$$

Thus, $\hat{\pi}_{1\Delta}$ is a set-valued decision rule that assigns a single treatment if only that treatment leads to an (estimated) expected clinically significant improvement on one or both outcomes and does not lead to a clinically significant loss in either outcome across all the treatment rules in $F(\hat{\pi}_{2\Delta})$ that a clinician might consider at the second stage. Alternatives to this definition of $\hat{\pi}_{1\Delta}$ are discussed in Section 5.

Remark 1

In addition to providing a set of recommended treatments it is useful to provide decision makers with information regarding outcomes which are likely to be realized under feasible regimes. At the second stage, estimates $(\hat{Q}_{2Y}(\mathbf{h}_2, 1), \hat{Q}_{2Z}(\mathbf{h}_2, 1))$ and $(\hat{Q}_{2Y}(\mathbf{h}_2, -1), \hat{Q}_{2Z}(\mathbf{h}_2, -1)$ should accompany $\hat{\pi}_{2\Delta}(\mathbf{h}_2)$. At the first stage, $\hat{\pi}_{1\Delta}(\mathbf{h}_1)$ should be accompanied by a plot of $\hat{Q}_{1Y}(\mathbf{h}_1, a_1, \tau_2)$ against $\hat{Q}_{1Z}(\mathbf{h}_1, a_1, \tau_2)$ across values of $\tau_2 \in F(\hat{\pi}_{2\Delta})$ with separate plotting symbols and colors for $a_1 = \pm 1$. Such a plot shows, for each potential first stage treatment, expected outcomes following feasible second stage treatment rules given current patient information as captured by $\mathbf{h}_1$. An example of such a plot is given in Figure 3.

3.1 Computation

Computing $\hat{\pi}_{1\Delta}(\mathbf{h}_1)$ requires solving a seemingly difficult enumeration problem. Nevertheless, exact computation of $\hat{\pi}_{1\Delta}(\mathbf{h}_1)$ is possible and (7) can be solved quickly when $p_{2,2}$ is small.

First, note that if τ_2 and τ_2' are decision rules at the second stage that agree on the observed data, that is, if $\tau_2(\mathbf{h}_{2,\cdot,i}) = \tau_2'(\mathbf{h}_{2,\cdot,i})$ for $i = 1, \ldots, n$, then $\hat{\psi}_{1Y}(\tau_2) = \hat{\psi}_{1Y}(\tau_2')$ and $\hat{\psi}_{1Z}(\tau_2) = \hat{\psi}_{1Z}(\tau_2')$. It follows that $\hat{\pi}_{1\Delta}(\mathbf{h}_1, \tau_2) = \hat{\pi}_{1\Delta}(\mathbf{h}_1, \tau_2') \ \forall \mathbf{h}_1 \in \mathbf{H}_1$. Thus, if we consider a finite subset $\tilde{F}(\hat{\pi}_{2\Delta})$ of $F(\hat{\pi}_{2\Delta})$ that contains one decision rule for each possible "labeling" of the stage 2 histories contained in the observed data, then we have

$$\widehat{\pi}_{1\Delta}(\mathbf{h}_1) = \bigcup_{\tau_2 \in F(\widehat{\pi}_{2\Delta})} \widehat{\pi}_{1\Delta}(\mathbf{h}_1, \tau_2) = \bigcup_{\tau_2 \in \tilde{F}(\widehat{\pi}_{2\Delta})} \widehat{\pi}_{1\Delta}(\mathbf{h}_1, \tau_2). \quad (8)$$

We use the term "labeling" by analogy with classification: each stage 2 history $\mathbf{h}_{2,\cdot,i}$ is given a binary "label" $\ell_i \in \{-1, 1\}$ by some τ_2. Rather than taking a union over the potentially uncountable $F(\hat{\pi}_{2\Delta})$ indexed by $\rho \in \mathbb{R}^{p2,2}$, we will instead enumerate the finite set of all feasible *labelings* of the observed data, place a corresponding τ_2 for each one into the set $\tilde{F}(\hat{\pi}_{2\Delta})$, and take the union over $\tilde{F}(\hat{\pi}_{2\Delta})$.

We say that a labeling $\ell_1, \ldots, \ell_n$ is *compatible* with a set-valued decision rule $\hat{\pi}_{2\Delta}$ if $\ell_i \in \hat{\pi}_{2\Delta}(\mathbf{h}_{2,\cdot,i})$, $i = 1, \ldots, n$, and *feasible* if it furthermore can be induced by a feasible decision rule $\tau_2 \in F(\hat{\pi}_{2\Delta})$. (Recall that in our terminology, feasible decision rules are compatible.) Equivalently, the labeling is feasible if it is both compatible with $\hat{\pi}_{2\Delta}$ and if the two sets $\{\mathbf{h}_{2,2,i} | \ell_i = 1\}$ and $\{\mathbf{h}_{2,2,i} | \ell_i = -1\}$ are *linearly separable* in $\mathbb{R}^{p2,2}$. Our algorithm for computing $\tilde{F}(\hat{\pi}_{2\Delta})$ works by specifying a linear mixed integer program with indicator constraints whose solutions correspond to the linearly separable labelings of the observed data that are compatible with $\hat{\pi}_{2\Delta}$.

First, we note that determining whether or not a given $\hat{\pi}_{2\Delta}$-compatible labeling $\ell_1, \ldots, \ell_n$ is feasible is equivalent to checking the following set of constraints:

$$\exists \psi_2 \text{ s.t. } \ell_i \mathbf{h}_{2,2,i}^\top \psi_2 \geq 1 \ \ \forall i \in 1, \ldots, n. \quad (9)$$

The constant 1 in the above inequalities is arbitrary since one can rescale both sides by any positive constant. Given a particular labeling, the existence of a ψ_2 that satisfies (9) can be proved or disproved in polynomial time using a linear program solver (see, e.g., Megiddo (1987) and references therein). The existence of such a ψ_2 implies a feasible τ_2 given by $\tau_2(\mathbf{h}_2) = \text{sgn}(\mathbf{h}_{2,2}^\top \psi_2)$ that produces the labeling $\ell_1, \ldots, \ell_n$ when applied to the stage 2 data.

To find all possible feasible labelings, we treat the ℓ_i as variables in an optimization, we formulate a linear mixed integer program with constraints given by

$$\begin{aligned} &\min_{\substack{\ell_1, \ell_2, \ldots, \ell_n \\ \psi_2 \in \mathbb{R}^{p_{2,2}}}} f(\ell_1, \ell_2, \ldots, \ell_n, \psi_2) \\ &\text{s.t.} \forall i \in 1, \ldots, n, \ \ \ell_i \in \hat{\pi}_{2\Delta}(\mathbf{h}_{2,\cdot,i}), \ \ \ell_i \mathbf{h}_{2,2,i}^\top \psi_2 \geq 1 \end{aligned}$$

and we find all unique feasible solutions. We present the feasibility problem as a minimization because it is the natural form for modern optimization software packages like CPLEX (www.ibm.com/software/integration/optimization/cplex-optimizer/), which are capable of handling the integer constraints on ℓ_i. Note that if we simply want to recover the feasible ℓ_i then the choice of f does not matter, and we may choose $f = 0$ for simplicity and efficiency in practice. CPLEX is capable of enumerating *all* feasible labelings efficiently (the examples considered in this paper take less than one minute to run on a laptop with 8GB DDR3 RAM and a 2.67GHz dual core processor). If we wanted to also recover the maximum margin separators for the feasible labelings, for example, we could use the quadratic objective $f = \|\psi_2\|^2$.

Let $F(\tilde{\pi}_{2\Delta})$ denote the collection of feasible decision rules defined by $\text{sgn}(\mathbf{h}_{2,2}^\top \psi_2)$ for each feasible ψ_2. Then for any $\mathbf{h}_1 \in \mathbb{R}^{p_1}$ we define $\hat{\pi}_{1\Delta}(\mathbf{h}_1)$ using (8). Note that $F(\hat{\pi}_{2\Delta})$ does not depend on the $\mathbf{h}_1$ and hence is computed only once for a given dataset.

4. CATIE

We now consider the application of our method to data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Schizophrenia study. The CATIE study was designed to compare sequences of antipsychotic drug treatments for the care of schizophrenia patients. The full study design is quite complex (Stroup et al., 2003); we will make several simplifications in order to more clearly illustrate the potential of our method. CATIE was an 18-month sequential randomized trial that was divided into two main phases of treatment. Upon entry into the study, most patients began "Phase 1," in which they were randomized uniformly to one of five possible treatments: olanzapine, risperidone, quetiapine, ziprasidone, or perphenazine. As they progressed through the study, patients were given the opportunity at each monthly visit to discontinue their Phase 1 treatment and begin "Phase 2" on a new treatment. The set of possible Phase 2 treatments depended on the reason for discontinuing Phase 1 treatment. If the Phase 1 treatment was deemed to produce unacceptable side-effects, they entered the *tolerability group* and their Phase 2 treatment was chosen randomly as follows: ziprasidone with probability 1/2, or uniformly randomly from the set {olanzapine, risperidone, quetiapine} with probability 1/2. If the Phase 1 treatment was deemed ineffective at reducing symptoms, they entered the *efficacy group* and

their Phase 2 treatment was chosen randomly as follows: clozapine with probability 1/2, or uniformly randomly from the set {olanzapine, risperidone, quetiapine} with probability 1/2.

Although CATIE was designed to compare several treatments within each stage, there are natural groupings at each stage that allow us to collapse the data in a meaningful way and consider only binary treatments. We can therefore directly apply our method as described. In the Phase 2 Tolerability group, it is natural to compare olanzapine against the other three drugs since it is known a priori to be efficacious (Breier et al., 2005), but is also known to cause significant weight gain as a side-effect. In the Phase 2 Efficacy group, it is natural to compare clozapine against the rest of the potential treatments, both because the randomization probabilities called for having 50% of patients in that group on clozapine, and because clozapine is substantively different from the other three drugs: it is known to be highly effective at controlling symptoms, but it is also known to have significant side-effects and its safe administration requires very close patient monitoring. In Phase 1, it is natural to compare perphenazine, the only typical antipsychotic, against the other four drugs which are atypical antipsychotics. (Typical-versus-atypical was an important comparison in CATIE.)

For our first outcome, which we denote *P*, we use the Positive and Negative Syndrome Scale (PANSS) which is a numerical representation of the level of psychotic symptoms experienced by a patient (Kay et al., 1987). A higher value of PANSS indicates more severe symptoms. PANSS is a well-established measure that we have used in previous work (Shortreed et al., 2011). Since having larger PANSS is worse, we use 100 minus the percentile of a patient's PANSS at the end of their time in the study. We use the distribution of PANSS at the beginning of the study as the reference distribution for the percentile.

For our second outcome, which we denote *B*, we use Body Mass Index (BMI), a measure of obesity. Weight gain is a clinically important side-effect of many antipsychotic drugs (Allison et al., 1999). Because in this population having a larger BMI is worse, we use 100 minus the percentile of a patient's BMI at the end of their time in the study. Again, we use the distribution of BMI at the beginning of the study as the reference distribution.

We transformed both outcomes into percentiles to match Lizotte et al. (2012); we also include an analysis using the raw BMI and PANSS scores in Web Appendix C. Regression diagnostics for both analyses as well as baseline distributions are given in Web Appendix D.

In all of our models, we include two baseline covariates. The first, TD, is an indicator of "tardive dyskinesia," a motor side-effect that can be caused by previous treatment. The second, EXACER, an indicator that the patient has been recently hospitalized, thus indicating an exacerbation of his or her condition. These do not interact with treatment in our models.

For our covariates $\mathbf{h}_2$ that interact with treatment, we choose the patients most recently recorded PANSS score percentile in our model for PANSS, and the most recently recorded BMI percentile in our model for BMI. These percentiles were shifted by −50 so that a patient at the median has $\mathbf{h}_2 = 0$. This was done so that in each model, the coefficient for the main effect of treatment can be directly interpreted as the treatment effect for a patient with median PANSS (resp. BMI). Treatments were coded −1, 1. For both outcomes we chose 5 percentile points as our indifference range, that so $\Delta_P = \Delta_B = 5$.

4.0.1 Phase 2 Tolerability

Web Appendix D gives the models estimated from the Phase 2 tolerability data. In summary, olanzapine appears to be beneficial if one considers the PANSS (*P*) outcome, but detrimental if one considers the BMI (*B*) outcome. This is evident in the center panel of

Figure 2, where we see that the predictions of (r_{2P}, r_{2B}) for all of the patient histories in our dataset fall in the lower-right region of the plot, where both treatments are recommended because they conflict with each other according to the two outcomes.

4.0.2 Phase 2 Efficacy

Web Appendix D gives the models estimated from the Phase 2 efficacy data. Clozapine appears to be beneficial if one considers the PANSS (P) outcome. Furthermore, there is weak evidence that clozapine is detrimental if one considers the BMI (B) outcome. This is evident in the rightmost panel of Figure 2, where the predictions of (r_{2P}, r_{2B}) for all of the patient histories are to the right of $\hat{r}_{2P} = \Delta_P$, indicating that clozapine is predicted to be the better choice for all patients when considering only the PANSS outcome. Furthermore, for most subjects, clozapine is not significantly worse than the other (aggregate) treatment according to BMI; thus, for most of the histories only clozapine (i.e., {1}) would be recommended. However, we found that for patients with a BMI covariate greater than about 24 (i.e., those among the top best 25 percent according to BMI), clozapine is predicted to perform clinically significantly worse according to the BMI outcome, and both treatments (i.e., {−1, 1}) would be recommended for these patients.

4.0.3 Phase 1

Recall that given any history $\mathbf{h}_1$ at Phase 1, our predicted values $(\hat{r}_{1P}, \hat{r}_{1B})$ for that history depend not only on the history itself but on the future decision rule that will be followed subsequently. For illustrative purposes, Web Appendix D gives a particular model estimated from the Phase 1 data assuming a particular feasible decision rule for Phase 2 chosen from the 61, 659 feasible Phase 2 decision rules enumerated by our algorithm. (The estimated coefficients would be different had we used a different Phase 2 decision rule.) For this particular future decision rule, perphenazine performs somewhat worse according to PANSS than the atypical antipsychotics, and somewhat better according to BMI.

Whereas for the Phase 2 analyses we showed plots of different $(\hat{r}_{2P}, \hat{r}_{2B})$ for different histories, for Phase 1, we will show different $(\hat{r}_{1P}, \hat{r}_{1B})$ for a *fixed* history at Phase 1 as we vary the Phase 2 decision rule. Recall that our treatment recommendation for Phase 1 is the union over all feasible future decision rules of the treatments recommended for each future decision rule. The leftmost panel in Figure 2 shows the possible values of $(\hat{r}_{1P}, \hat{r}_{1B})$; for some future decision rules only treatment −1 is recommended, but for others the set {−1, 1} is recommended. Taking the union, we recommend the set {−1, 1} for this history at Phase 1. Figure 3 shows, for a fixed first stage history $\mathbf{h}_1$, a plot of $\hat{Q}_{1B}(\mathbf{h}_1, a_1, \tau_2)$ against $\hat{Q}_{1P}(\mathbf{h}_1, a_1, \tau_2)$ across all $\tau_2 \in \tilde{F}(\hat{\pi}_{2\Delta})$ for a single subject in the CATIE data. Note, that while there are 61,659 polices in $F(\pi_{2\Delta})$ many of these yield similar predicted values for $\hat{Q}_{1B}(\mathbf{h}_1, a_1, \tau_2)$ and $\hat{Q}_{1P}(\mathbf{h}_1, a_1, \tau_2)$. This display suggests that a patient presenting with $\mathbf{H}_1 = \mathbf{h}_1$, choosing perphenazine (PERP) is associated with better expected outcomes on BMI but worse on PANSS under feasible second stage rules.

5. Discussion

We proposed set-valued dynamic treatment regimes as a method for adapting treatment recommendations to the evolving health status of a patient in the presence of competing outcomes. Our proposed methodology deals with the reality that there is typically no universally best treatment for chronic illnesses like depression or schizophrenia by identifying when a trade-off between efficacy and side-effects must be made. Although computation of the set-valued dynamic treatment regimes requires solving a difficult enumeration problem, we offered an efficient algorithm that uses existing linear mixed integer programming software.

Our proposed methodology avoids the construction of composite outcomes, a process which may be undesirable: constructing a composite outcome requires combining outcomes that are on different scales, the 'optimal trade-off' between two (or more) outcomes is likely to be patient-specific, evolving over time, and the assumption that a linear trade-off is sufficient to describe all possible patient preferences may be unrealistic.

There are a number of directions in which this work can be extended. Web Appendix A provides an extension to the case with two decision points but an arbitrary number of treatment choices available at each stage. Interestingly, our enumeration problem is closely related to *transductive learning*, a classification problem setting where only a subset of the available training data is labeled, and the task is to predict labels at the unlabeled points in the training data. By finding a minimum-norm solution for ψ_2 subject to our constraints, we could produce the transductive labeling that induces the maximum margin linear separator. Our algorithm would then correspond to a linear separable transductive support vector machine (SVM) (Cortes and Vapnik, 1995). This observation leads to a possible criterion for evaluating feasible decision rules: we hypothesize that the greater the induced margin, the more "attractive" the decision rule, because large-margin decision rules avoid giving very similar patients different treatments. If the number of feasible future decision rules becomes impractically large, we may wish to keep only the most "separable" ones when computing the union at the first stage. We are currently pursuing this line of research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge invaluable comments and criticisms from the Editor, Associate Editor, and two anonymous referees. Eric Laber acknowledges support from NIH grant P01 CA142538. Data used in the preparation of this article were obtained from the limited access datasets distributed from the NIH-supported "Clinical Antipsychotic Trials of Intervention Effectiveness in Schizophrenia" (CATIE-Sz). The study was supported by NIMH Contract #N01MH90001 to the University of North Carolina at Chapel Hill. The ClinicalTrials.gov identifier is NCT00014001. This manuscript reflects the views of the authors and may not reflect the opinions or views of the CATIE-Sz Study Investigators or the NIH.

References

Allison DB, Mentore JL, Heo M, Chandler LP, Cappelleri JC, Infante MC, Weiden PJ. Antipsychotic-induced weight gain: A comprehensive research synthesis. American Journal of Psychiatry. Nov. 1999 156:1686–1696. [PubMed: 10553730]

Bellman, R. Dynamic Programming. Princeton University Press; 1957.

Blatt, D.; Murphy, SA.; Zhu, J. Technical Report 04-63. The Methodology Center, Penn. State University; 2004. A-learning for approximate planning.

Breier A, Berg PH, Thakore JH, Naber D, Gattaz WF, Cavazzoni P, Walker DJ, Roychowdhury SM, Kane JM. Olanzapine versus ziprasidone: Results of a 28-week double-blind study in patients with schizophrenia. American Journal of Psychiatry. 2005; 162:1879–1887. [PubMed: 16199834]

Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20:273–297.

Friedman, LM.; Furberg, CD.; DeMets, DL. Fundamentals of clinical trials. Springer; 2010.

Henderson R, Ansell P, Alshibani D. Regret-regression for optimal dynamic treatment regimes. Biometrics. 2010; 66:1192–1201. [PubMed: 20002404]

Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophrenia Bulletin. 1987; 13(2):261–276. [PubMed: 3616518]

Kinter, ET. Identifying Treatment Preferences of Patients with Schizophrenia in Germany: An Application of Patient-centered Care. ProQuest; 2009.

Laber EB, Lizotte DJ, Qian M, Murphy SA. Statistical inference in dynamic treatment regimes. Preprint, arXiv:1006.5831v1. 2011

Lizotte DJ, Bowling M, Murphy S. Linear fitted-Q iteration with multiple reward functions. Journal of Machine Learning Research. 2012 Accepted.

Megiddo N. On the complexity of linear programming. Advances in economic theory. 1987:225–268.

Moodie EEM, Richardson TS, Stephens DA. Demystifying optimal dynamic treatment regimes. Biometrics. 2007; 63(2):447–455. [PubMed: 17688497]

Murphy SA. Optimal dynamic treatment regimes (with discussion). Journal of the Royal Statistical Society: Series B. 2003; 58:331–366.

Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano G, Waxmonsky J, Yu J, Murphy SA. Q-learning: A data analysis method for constructing adaptive interventions. Technical Report. 2010

Nahum-Shani I, Qian M, Almirall D, Pelham WE, Gnagy B, Fabiano G, Waxmonsky J, Yu J, Murphy SA. Experimental design and primary data analysis methods for comparing adaptive interventions. To appear, Psychological Methods. 2012

Orellana L, Rotnitzky A, Robins J. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: Main content. Int Jrn of Biostatistics. 2010; 6(2)

Robins, J. Optimal structural nested models for optimal sequential decisions. Proceedings of the second seattle Symposium in Biostatistics; Springer; 2004. p. 189-326.

Schulte P, Tsiatis A, Laber E, Davidian M. Q-and a-learning methods for estimating optimal dynamic treatment regimes. Statistical Science. 2012 In Press.

Shortreed S, Laber EB, Lizotte DJ, Stroup TS, Pineau J, Murphy SA. Informing sequential clinical decision-making through reinforcement learning: an empirical study. Machine Learning. 2011; 84(1–2):109–136. [PubMed: 21799585]

Song R, Wang W, Zeng D, Kosorok MR. Penalized q-learning for dynamic treatment regimes. Pre-Print, arXiv:1108.5338v1. 2012

Strauss GP, Robinson BM, Waltz JA, Frank MJ, Kasanova Z, Herbener ES, Gold JM. Patients with schizophrenia demonstrate inconsistent preference judgments for affective and nonaffective stimuli. Schizophrenia bulletin. 2011; 37(6):1295–1304. [PubMed: 20484522]

Strecher VJ, Shiffman S, West R. Moderators and mediators of a web-based computer-tailored smoking cessation program among nicotine patch users. Nicotine & tobacco research. 2006; 8(S. 1):S95. [PubMed: 17491176]

Stroup TS, McEvoy JP, Swartz MS, Byerly MJ, Glick ID, Canive JM, McGee MF, Simpson GM, Stevens MC, Lieberman JA. The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) project: Schizophrenia trial design and protocol development. Schizophrenia Bulletin. 2003; 29(1):15–31. [PubMed: 12908658]

Watkins CJCH, Dayan P. Q-learning. Machine Learning. 1992; 8:279–292.

Zhang B, Tsiatis A, Laber E, Davidian M. A robust method for estimating optimal treatment regimes. Biometrics. 2012 To appear.

Zhao Y, Zeng D, Rush JA, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. 2012 Submitted.

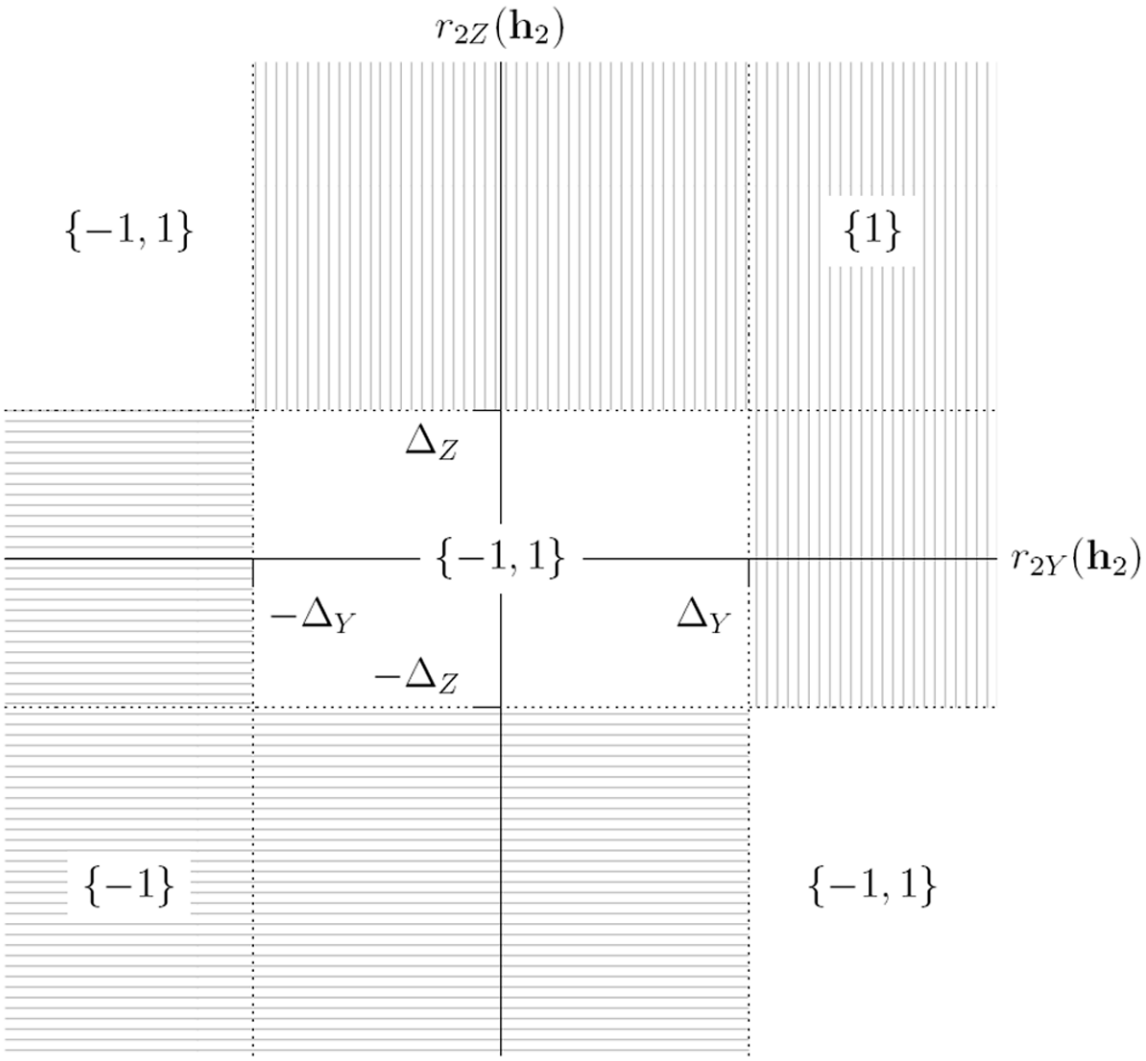

Figure 1.

Diagram showing how the output of $\pi_{2\Delta}^{\text{Ideal}}(\mathbf{h}_2)$ depends on Δ_Y and Δ_Z, and on the location of the point ($r_{2Y}(\mathbf{h}_2)$, $r_{2Z}(\mathbf{h}_2)$).

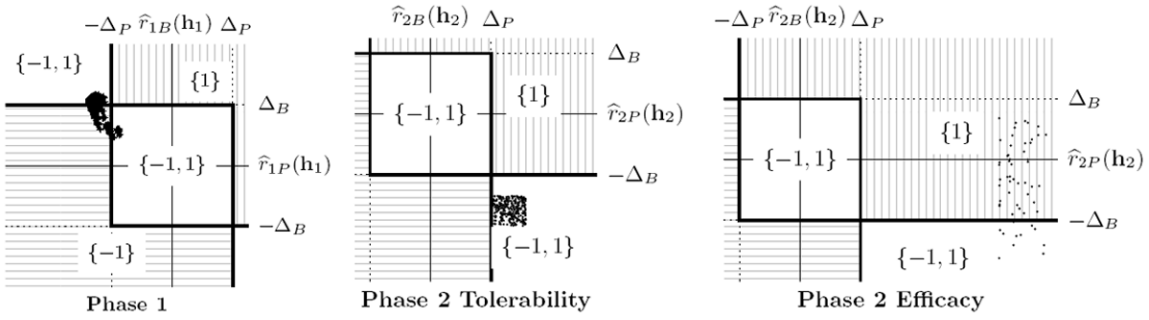

Figure 2.

Left: Diagram showing how the output of $\pi_{1\widehat{\Delta}}(\mathbf{h}_1)$ depends on Δ_P (clinically significant difference in PANSS) and Δ_B (clinically significant difference in BMI), and on the joint treatment effect, at Phase 1. The cloud of points shows the possible joint treatment effects that can be realized by a *single patient* with history (panss = −25.5, bmi = −15.6) if the patient follows some feasible decision rule at Phase 2. That is, each point is associated with a different choice of Phase 2 decision rule. Note that for some future decision rules, the point lies in the {−1, 1} region, and for others it lies in the {−1} region; taking the union we have $\pi_{1\widehat{\Delta}}(\mathbf{h}_1) = \{-1, 1\}$ for this patient. **Center:** Diagram showing how the output of $\pi_{2\widehat{\Delta}}(\mathbf{h}_2)$ depends on Δ_P and Δ_B, and on the location of the point $(r_{2\widehat{P}}(\mathbf{h}_2), r_{2\widehat{B}}(\mathbf{h}_2))$ for *all patients* in the Phase 2 Tolerability group. Each plotted point shows the estimated joint treatment effect for a *different patient* in the dataset. Since Phase 2 is the last phase, there are no future decision rules to consider and each history is associated with a unique joint treatment effect. **Right:** Analogous plot for the Phase 2 Efficacy group.

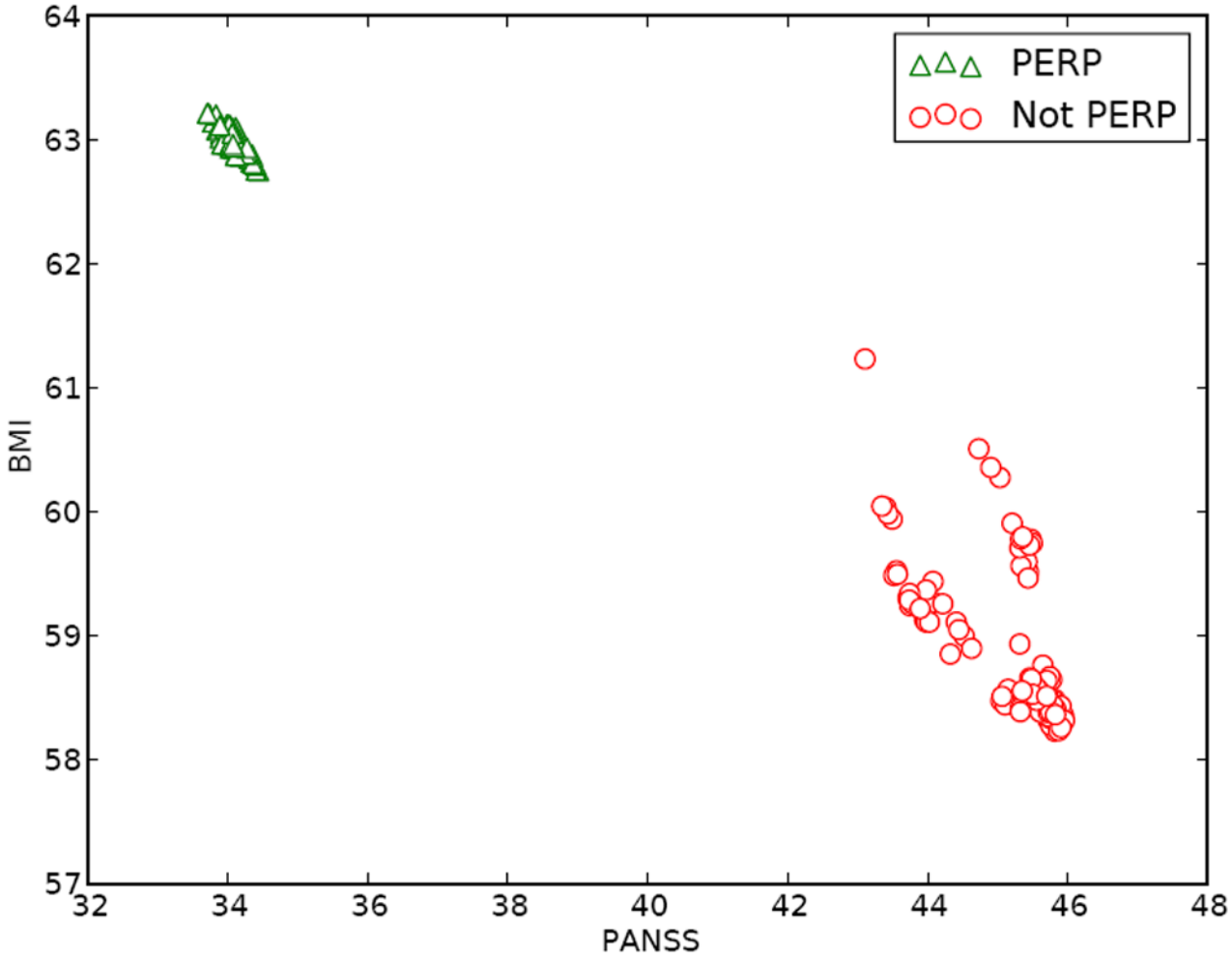

Figure 3.

$\hat{Q}_{1B}(\mathbf{h}_1, a_1, \tau_2)$ against $\hat{Q}_{1P}(\mathbf{h}_1, a_1, \tau_2)$ across $\tau_2 \in F(\tilde{\pi}_{2\Delta})$ for a single patient with history (panss = −25.5, bmi = −15.6) in the CATIE data. Note, that while there are 61,659 polices in $F(\tilde{\pi}_{2\Delta})$ many of these yield similar predicted values for $\hat{Q}_{1B}(\mathbf{h}_1, a_1, \tau_2)$ and $\hat{Q}_{1P}(\mathbf{h}_1, a_1, \tau_2)$; we have plotted a random subset to make individual points more clearly visible. This display suggests that a patient presenting with $\mathbf{H}_1 = \mathbf{h}_1$, choosing perphenazine (PERP) is associated with better expected outcomes on BMI but worse on PANSS under feasible second stage rules.

Table 1

Illustrative example for competing outcomes generically called ‘side-effects’ and ‘efficacy’ and no patient covariates. In both cases the set {−1, 1} should be recommended at the first stage. Subject A illustrates the impact of preference evolution. If Subject A is initially concerned only with efficacy then they will choose treatment 1 at the first stage. However, if at the time of the second decision Subject A is concerned with side-effects, having initially chosen treatment 1 they are only left with poor choices. Subject B illustrates a potential problem with applying Q-learning separately to each outcome and then checking for agreement. Both Q-learning applied to efficacy and side-effects recommend treatment 1 at the first stage for Subject B. However, the Q-learning algorithm applied to efficacy assumes treatment -1 will be chosen at the second stage, and the Q-learning algorithm applied to side-effects assumes treatment 1 will be applied at the second stage. Yet, for Subject B applying treatment 1 at the first stage leads to extreme and potentially undesirable trade-offs at the second stage.

		Subject A		Subject B	
Stage 1 Txt	**Stage 2 Txt**	**Side-effects**	**Efficacy**	**Side-effects**	**Efficacy**
1	1	Very poor	Very good	Very good	Very poor
1	-1	Poor	Poor	Poor	Very good
-1	1	Good	Good	Good	Good
-1	-1	Very good	Poor	Fair	Very poor