Workshop on Personalized Medicine and Dynamic Treatment Regimes

Marie Davidian, Butch Tsiatis, Eric B. Laber, and Michael Kosorok

IMPACT Symposium, November 1, 2012

Workshop Outline

Introduction to Personalized Medicine and Dynamic Treatment Regimes

Estimation of Optimal Dynamic Treatment Regimes for a Single Decision

Estimation of Optimal Dynamic Treatment Regimes for Multiple Decisions

Advanced Topics in Personalized Medicine and Dynamic Treatment Regimes

Workshop Outline

Introduction to Personalized Medicine and Dynamic Treatment Regimes

Estimation of Optimal Dynamic Treatment Regimes for a Single Decision

Estimation of Optimal Dynamic Treatment Regimes for Multiple Decisions

Advanced Topics in Personalized Medicine and Dynamic Treatment Regimes

Outline

- Personalized Medicine
- Clinical Decision-Making and Dynamic Treatment Regimes
- Optimal Dynamic Treatment Regime
- Discovery (Estimation) of Optimal Dynamic Treatment Regimes
- Clinical Trials for Discovery of Dynamic Treatment Regimes
- Roadmap for the Workshop

What is Personalized Medicine?

"*The right treatment for the right patient (at the right time)*"

What is Personalized Medicine?

In general: One size does *not* fit all

- Multiple *treatment options* may be available
- Patient *heterogeneity*
  - *Across patients:* What works for one patient may not work for another
  - *Within patients:* What works now may not work later

Premise: Use *information* on a patient's characteristics to determine which treatment option s/he should receive (and when. . .)

- Genetic/genomic, demographic,. . .
- Physiologic/clinical measures, medical history,. . .

Popular Perspective on Personalized Medicine

Subgroup identification/targeted treatment:

- Are there subgroups of patients who are *more likely* to do better on one particular treatment than on another?
- Can a treatment be developed that *targets* a subgroup that is very likely to benefit from that treatment?
- Can *biomarkers* be developed to identify such patients?

Focus: Treating and targeting treatment for *subgroups* of the population

Another Perspective on Personalized Medicine

Can we determine how to treat the entire population of patients?

- *Given information on patient's characteristics*, can we determine the treatment from among the available options most likely to benefit him/her?
- And by doing so determine how best to treat the *population*?
- *This is the perspective we will take in this workshop*

Clinical Decision-Making

Clinical practice: Clinicians make (a series of) *treatment decision(s)* over the course of a patient's disease or disorder

- *Fixed schedule*
- *Milestone* in the disease process
- *Event* necessitating a decision

Clinical decision-making: *Clinical judgment*

- Synthesize all *information* on a patient up to the point of a decision to determine next treatment action
- *Goal:* "*Individualize*" the decision to the patient
- Can this be *formalized* and made *evidence-based*?

Dynamic Treatment Regime

Operationalizing personalized medicine: At any *decision point*

- Construct a *rule* that takes as *input* the *available information* on the patient to that point and dictates the next treatment from among the *possible, feasible options*
- Rule(s) must be developed based on *evidence*, i.e., *data*

Dynamic treatment regime: A set of formal *rules*, each corresponding to a *decision point*

- Each rule dictates the *treatment action* to be taken at that point as a *function of accrued information* on the patient
- Together, the rules determine an *algorithm* for treating any patient, referred to collectively as a *dynamic treatment regime*

Dynamic Treatment Regime

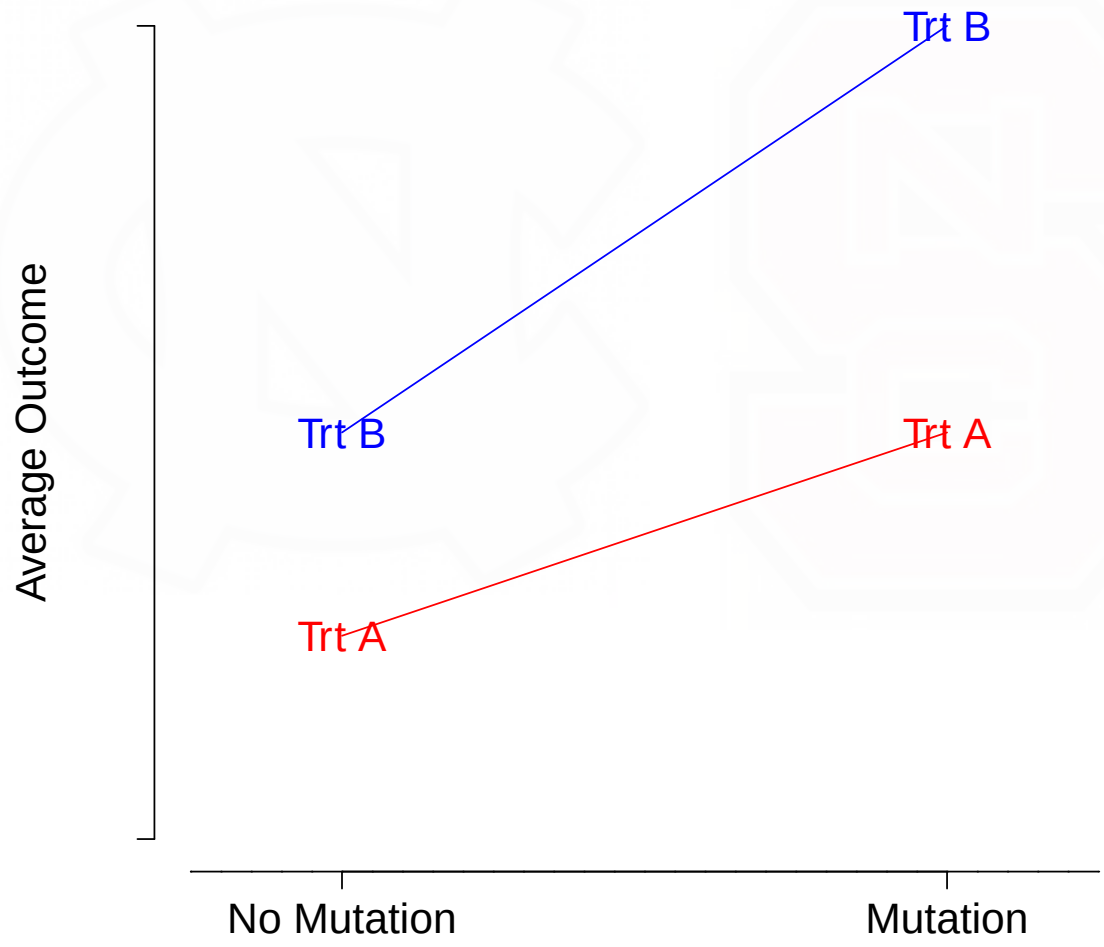
Assume: There is a *clinical outcome* by which treatment benefit can be assessed

- Survival time, CD4 count, indicator of no myocardial infarction within 30 days, ...
- *Larger outcomes* are *better*

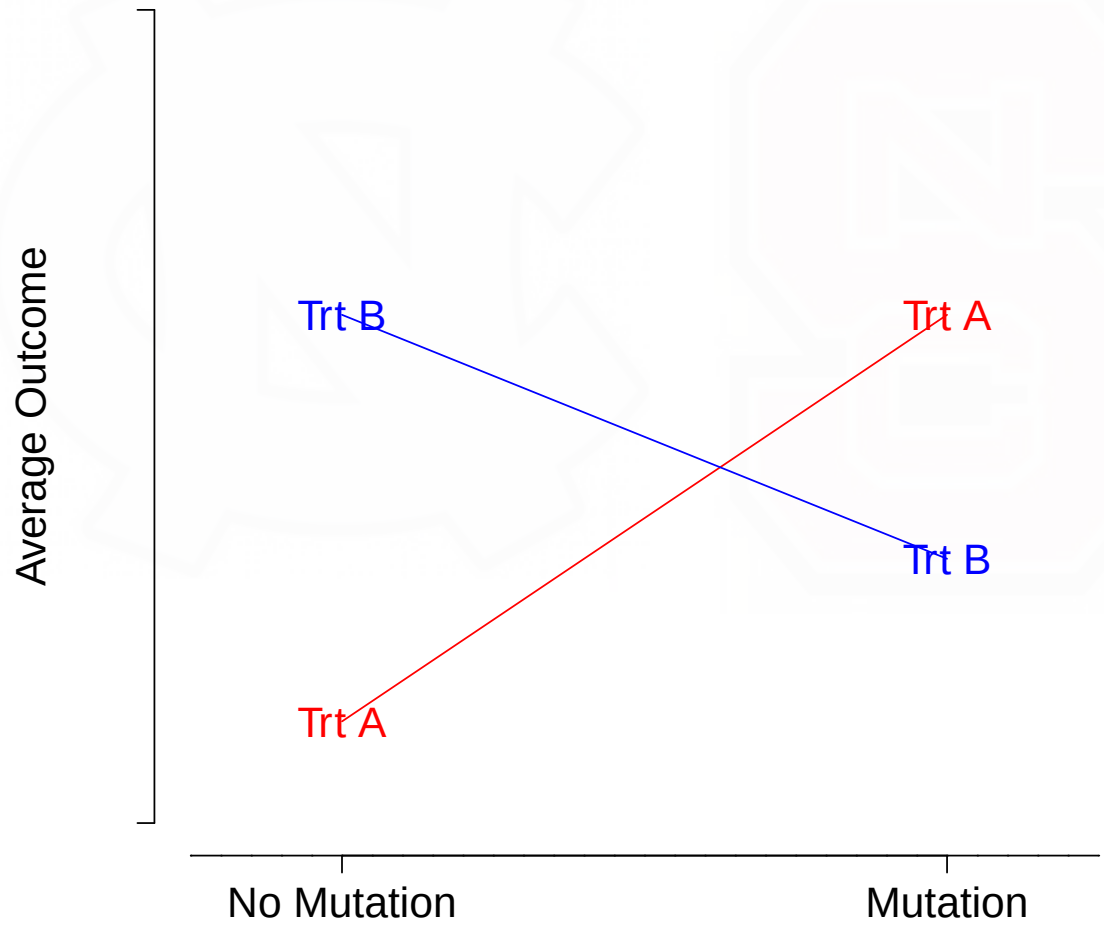
Intuitively: Rules should depend on *characteristics* (*variables*, *covariates*) that exhibit a *qualitative interaction* with treatment

- "*Tailoring variables*"

Tailoring Variables



Tailoring Variables



Single Decision Point

Simple example: Which treatment to give patients who present with *primary operable breast cancer*?

- *Options:* L-phenylalnanine mustard and 5-flourouracil (PF) or PF + tamoxifen (PFT)
- *Data:* $\sim$ 1,300 patients in a National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trial (Gail and Simon, 1985)
- *Available information:* age (years), progesterone receptor (PR) level (fmol)
- *Outcome:* Disease-free survival to three years

Single Decision Point

Gail and Simon rule/regime:

- If age < 50 and PR < 10 fmol $\Rightarrow$ PF (1); else $\Rightarrow$ PFT (0)
- *Mathematically:* The formal rule is

$$d(\text{age,PR}) = I(\text{age} < 50 \text{ and PR } < 10)$$

Alternatively: Rules of form

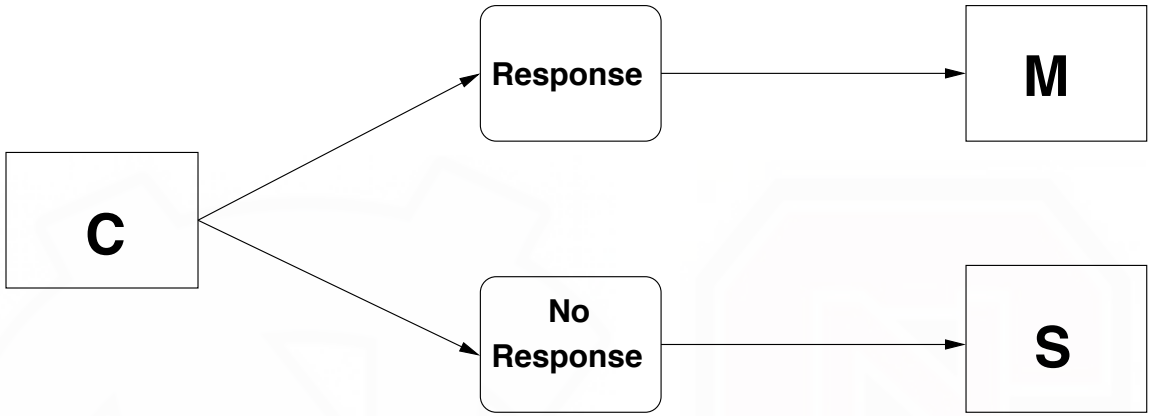
$$d(\text{age,PR}) = I(\text{age} > 60 - 8.7\log(\text{PR}))$$

Multiple Decision Points

Cancer treatment: *Two decision points*

- *Decision point 1:* Induction chemotherapy
- *Decision point 2:* Maintenance/intensification treatment (responders), Salvage chemotherapy (nonresponders)
- *Outcome:* Survival time

Multiple Decision Points



- *At presentation:* Information x_1; *accrued information* $h_1 = x_1$
- *Decision point 1:* Three *options* $\{c_1, c_2, c_3\}$; *rule 1:* $d_1(h_1) \Rightarrow d_1 : h_1 \rightarrow \{c_1, c_2, c_3\}$
- *Between decisions 1 and 2:* Collect *additional information* x_2, including *responder status*
- *Accrued information* $h_2 = \{x_1,$ chemotherapy at decision 1, $x_2\}$
- *Decision point 2:* Four *options* $\{m_1, m_2, s_1, s_2\}$; *rule 2:* $d_2(h_2) \Rightarrow d_2 : h_2 \rightarrow \{m_1, m_2\}$ (responders), $d_2 : h_2 \rightarrow \{s_1, s_2\}$ (nonresponders)

Summary

Single decision: 1 decision point

- *Information* x
- *Decision rule* $d(x)$, $d : x \to \mathcal{A}$ = set of *treatment options* a
- *Treatment regime:* d

Multiple decisions: K decision points

- Initial *information* x_1, *intermediate information* x_k between decisions $k-1$ and k, $k = 2, \ldots, K$
- Set of *treatment options* at decision k $a_k \in \mathcal{A}_k$
- *Accrued information* $h_1 = x_1$,
  $h_k = \{x_1, a_1, x_2, a_2, \ldots, x_{k-1}, a_{k-1}, x_k\}$, $k = 2, \ldots, K$
- *Decision rules* $d_1(h_1), d_2(h_2), \ldots, d_K(h_K)$, $d_k : h_k \to \mathcal{A}_k$
- *Dynamic treatment regime* $d = (d_1, d_2, \ldots, d_K)$

Considerations

Realistically: *High-dimensional information* x_k, $k = 1, \ldots, K$

- Must construct rules that *distill* this information
- Must identify the (likely very small) *subset* that are good *tailoring variables*

Furthermore: *Many* possible regimes d

- $\mathcal{D}$ = class of all possible dynamic treatment regimes
- Can we find the "*best*" set of rules; i.e., the "*best*" dynamic treatment regime in $\mathcal{D}$?

Optimal Dynamic Treatment Regime

How do we define "best"?

- If an individual patient *were to receive treatment* according to the set of rules $d_1, \ldots, d_K$, that is, according to regime $d = (d_1, \ldots, d_K)$, his/her *expected outcome* would be *as large as possible given the information available on him/her*
- If all patients in the *population* were to receive treatment according to regime d, the *expected (average) outcome* for the population would be *as large as possible given the information available*
- Can we *formalize* this?

Potential Outcomes

Single decision: Possible treatment options $a \in \mathcal{A}$

- For a *randomly chosen patient* from the population, define the *random variable* $Y^*(a)$ = the outcome the patient *would experience* if s/he *were to receive* treatment option a
- "*Potential outcome*"
- E.g., if $\mathcal{A} = \{0, 1\}$ (two possible treatment options), $Y^*(1)$= the outcome a patient would have if s/he were given treatment 1, and similarly for $Y^*(0)$
- Define $Y^*(d)$ = the outcome a patient would have if s/he received treatment *according to a regime* $d \in \mathcal{D}$
- E.g., if $\mathcal{A} = \{0, 1\}$ and the patient has information X

$$Y^*(d) = Y^*(1)d(X) + Y^*(0)\{1 - d(X)\}$$

Optimal Dynamic Treatment Regime

Single decision, continued:

- $E\{Y^*(d)|X = x\}$ is the *expected outcome* for a patient with *information* x if s/he were to receive treatment according to regime $d \in \mathcal{D}$
- $E\{Y^*(d)\} = E[\,E\{Y^*(d)|X\}\,]$ is the expected (average) outcome for the *population* if *all patients* were to receive treatment according to regime $d \in \mathcal{D}$

Optimal regime: d^{opt} is a regime in $\mathcal{D}$ such that

- $E\{Y^*(d)|X = x\} \leq E\{Y^*(d^{opt})|X = x\}$ for all $d \in \mathcal{D}$ and all values of x
- And thus $E\{Y^*(d)\} \leq E\{Y^*(d^{opt})\}$ for all $d \in \mathcal{D}$

Optimal Dynamic Treatment Regime

Multiple decisions: Same idea, only more *complicated*

- *Initial information* X_1
- *Potential outcomes* under a regime $d \in \mathcal{D}$

$$X_2^*(d), \ldots, X_K^*(d), Y^*(d)$$

- $E\{Y^*(d)|X_1 = x_1\} \leq E\{Y^*(d^{opt})|X_1 = x_1\}$ for all $d \in \mathcal{D}$ and values of x_1
- And thus $E\{Y^*(d)\} \leq E\{Y^*(d^{opt})\}$ for all $d \in \mathcal{D}$

Important Philosophical Point

Distinguish between:

- The "*best*" treatment for a patient
- The "*best*" treatment for a patient *given the information available*

Best treatment for a patient: Option $a^{best} \in \mathcal{A}$ corresponding to the *largest* $Y^*(a)$ for that patient

Best treatment given the information available:

- We *cannot* hope to determine a^{best} because we can never see *all* the potential outcomes on a given patient
- What we *can* hope to do is to make the *optimal decision* given the *information available* $\Rightarrow$ find d^{opt} and make $E\{Y^*(d^{opt})|X = x\}$ as large as possible

Discovery (Estimation) of Optimal Dynamic Treatment Regimes

Result: This perspective on *personalized medicine* boils down to *discovery of* optimal dynamic treatment regimes based *data*

- *Existing data* from *observational studies* (e.g., registries), previously conducted *clinical trials*
- *Prospectively collected data* from *clinical trials* designed specifically for this purpose (coming up)

Discovery (Estimation) of Optimal Dynamic Treatment Regimes

Single decision: *Data* (X_i, A_i, Y_i), $i = 1, \ldots, n$

- n subjects indexed by i
- $X_i =$ *information* observed on subject i
- $A_i =$ *observed treatment* actually received by subject i
- $Y_i =$ *observed outcome* for subject i
- *Goal:* Under suitable *assumptions*, *estimate* $d^{opt}(x)$ using these data

Discovery (Estimation) of Optimal Dynamic Treatment Regimes

Multiple decisions: Data

$$(X_{1i}, A_{1i}, X_{2i}, A_{2i}, \ldots, X_{(K-1)i}, A_{(K-1)i}, X_{Ki}, Y_i), \quad i = 1, \ldots, n$$

- X_{1i} = *Initial information* observed on subject i
- X_{ki}, $k = 2, \ldots, K$ = *intermediate information* between decisions $k-1$ and k on subject i
- A_{ki}, $k = 1, \ldots, K$ = *observed treatment* actually received by subject i at decision k
- Y_i = *observed outcome* for subject i; can be *ascertained after* decision K or can be a *function* of $X_{2i}, \ldots, X_{Ki}$
- *Goal:* Under suitable *assumptions*, *estimate* $d^{opt}(x)$ using these data

Discovery (Estimation) of Optimal Dynamic Treatment Regimes

Challenges:

1. The optimal dynamic treatment regime d^{opt} is defined in terms of *potential outcomes* (not the observed data)
2. Were all possibly useful *tailoring variables* that clinicians used in the study at each decision point *recorded* in the data?
3. This sounds *hard*; does it really have to be?

Discovery (Estimation) of Optimal Dynamic Treatment Regimes

Challenge 1: d^{opt} is defined in terms of *potential outcomes*

- Need to be able to express the definition of d^{opt} equivalently in terms of the *data*
- Possible under certain *assumptions*
- *Butch* will demonstrate in the *single decision* case
- Also possible in the *multiple decision* case (but harder)

Discovery (Estimation) of Optimal Dynamic Treatment Regimes

Challenge 3: Can't we just *piece together results* from several studies to figure out the optimal regime?

- Study comparing *induction chemotherapies* based on *response*
- Study comparing *maintenance therapies* based on *survival time* among *responders* to induction therapy
- Study comparing *salvage therapies* based on *survival time* among *nonresponders*
- Wouldn't the regime that uses the "*best*" option in each study have to have the "*best*" average outcome?
- *Delayed effects:* The induction therapy with the highest proportion of *responders* might have *other effects* that render subsequent treatments less effective in regard to *survival*
- *Result:* Must consider the *entire sequence* of decisions

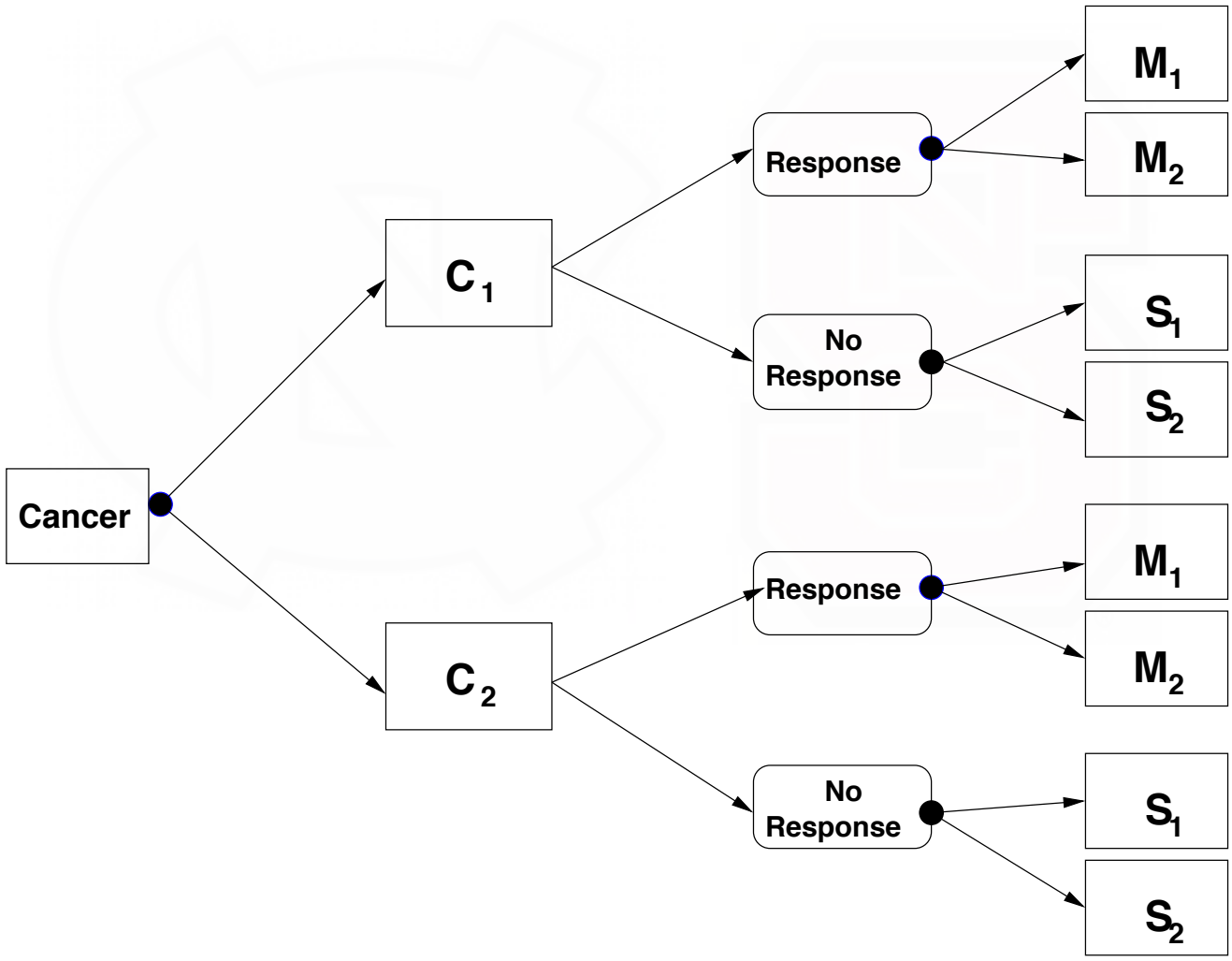
Clinical Trials for Discovery of Dynamic Treatment Regimes

Challenge 2: Conduct a *clinical trial* specifically designed for estimation of optimal dynamic treatment regimes

- *SMART:* Sequential Multiple Assignment Randomized Trial
- *Randomize* subjects to the *treatment options* at *each decision point*
- Collect *extensive, detailed information* initially and intermediate to decision points on possible *tailoring variables*

Later: *Eric and Michael* will have more to say on both of these issues

Clinical Trials for Discovery of Dynamic Treatment Regimes



Roadmap for Today

8:40 am – 9:30 am	*Butch* will discuss estimation of optimal treatment regimes for the *single decision* setting
9:30 am – 9:45 am	*Break*
9:45 am – 10:35 am	*Eric* will discuss estimation of optimal dynamic treatment regimes for the *multiple decision* setting, SMART studies
10:35 am – 10:50 am	*Break*
10:50 am – 11:30 am	*Michael* will provide an overview of some more *advanced topics*, including approaches for handling *high dimensional information* and *censored outcomes*, challenges associated with *making inference*, *open problems*

Workshop Outline

Introduction to Personalized Medicine and Dynamic Treatment Regimes

Estimation of Optimal Dynamic Treatment Regimes for a Single Decision

Estimation of Optimal Dynamic Treatment Regimes for Multiple Decisions

Advanced Topics in Personalized Medicine and Dynamic Treatment Regimes

Optimal Regime

Assume: *Large* outcomes are *good*

The optimal regime:

- The *regime* that, if followed by all patients in the population, yields the *largest outcome on average*

Goal: Given *data*, (*evidence*) from a clinical trial or observational study, *estimate* the *optimal regime* satisfying this definition

- *For simplicity*: Consider regimes involving a *single decision*/*rule*

Statistical Framework

Simplest setting: A *single decision* with *two* treatment options

Observed data: (Y_i, X_i, A_i), $i = 1, \ldots, n$, iid

- Y_i outcome, X_i baseline covariates, $A_i = 0, 1$ treatment received

Treatment regime: A single *rule*

- A function $d : X \to \{0, 1\}$

Application

Simple example: How to treat patients with primary operable *breast cancer* with positive nodes (a *single* decision point)?

- *Options*: L-phenylalanine mustard and 5-fluorouracil (PF) or PF + tamoxifen (PFT)
- *Data* from $\sim$ 1,300 patients in a National Surgical Adjuvant Breast and Bowel Project (NSABP) clinical trial (Gail and Simon, 1985)
- *Information*: age (years), progesterone receptor level (PR; fmol)

Application

Gail and Simon rule:

- If age $<$ 50 and PR $<$ 10 fmol $\Longrightarrow$ PF (1); otherwise $\Longrightarrow$ PFT (0)
- *Mathematically*, the rule is

$$d(\text{age}, \text{PR}) = I(\text{age} < 50 \text{ and } \text{PR} < 10)$$

- The *treatment regime* uses this rule to determine treatment

Statistical Framework

- *Even simpler example*: $d(X) = I(\text{age} \leq 50)$
- $d \in \mathcal{D}$, the class of *all* regimes
- *Optimal regime*: If followed by *all patients* in the population, would lead to *largest average outcome* among all regimes in $\mathcal{D}$

Potential Outcomes

Formalize: We can hypothesize *potential outcomes*

- $Y^*(1)$ = outcome that would be achieved if patient were to receive 1; $Y^*(0)$ defined similarly
- We *observe* $Y = Y^*(1)A + Y^*(0)(1 - A)$
- $\Longrightarrow E\{Y^*(1)\}$ is the *average outcome* if *all patients* in the population were to receive 1; and similarly for $E\{Y^*(0)\}$

Potential Outcomes

No unmeasured confounders: Assume that

$$Y^*(0), Y^*(1) \perp\!\!\!\perp A|X$$

- X contains all information used to assign treatments in the data
- Automatically satisfied for data from a *randomized trial*
- Standard but *unverifiable* assumption for *observational studies*

Potential Outcomes

- Implies that

$$\begin{aligned} E\{Y^*(1)\} &= E[E\{Y^*(1)|X\}] \\ &= E[E\{Y^*(1)|A=1,X\}] \\ &= E\{\,E(Y|A=1,X)\,\} \end{aligned}$$

and similarly for $E\{Y^*(0)\}$

Optimal Regime

Potential outcome for a regime:

- For any $d \in \mathcal{D}$, define $Y^*(d)$ to be the potential outcome for an arbitrary individual in our population if, possibly contrary to fact, he/she was assigned treatment in accordance to treatment regime d; that is,

$$Y^*(d) = Y^*(1)d(X) + Y^*(0)\{1 - d(X)\} \qquad (1)$$

- $E\{Y^*(d)\}$ is the mean response of a population all treated according to the regime d

Optimal Regime

- *Optimal regime*: Leads to *largest* $E\{Y^*(d)\}$ among all $d \in \mathcal{D}$; i.e.,

$$d^{opt} = \arg\max_{d \in \mathcal{D}} E\{Y^*(d)\}$$

- (1) implies that

$$\begin{aligned} E\{Y^*(d)\} &= E[E\{Y^*(d)|X\}] = E\Big[E\{Y^*(1)|X\}d(X) \\ &\quad + E\{Y^*(0)|X\}\{1-d(X)\}\Big] \\ &= E\Big[E(Y|A=1,X)d(X) + E(Y|A=0,X)\{1-d(X)\}\Big] \\ &= E[\mu(1,X)d(X) + \mu(0,X)\{1-d(X)\}], \end{aligned}$$

where $E(Y|A,X) = \mu(A,X)$

Optimal Regime

- $d^{opt}(x) = \arg\max_{a=\{0,1\}} E\{Y^*(a)|X = x\}$
- *Thus* $d^{opt}(X) = I[E\{Y^*(1)|X\} > E\{Y^*(0)|X\}] = I\{\mu(1, X) > \mu(0, X)\}$
- *Result*: If $E(Y|A, X) = \mu(A, X)$ were *known*, we could find d^{opt}

Estimating the Optimal Regime

Problem: $E(Y|A,X)$ is *not known*

- *Posit a model* $\mu(A,X;\beta)$ for $E(Y|A,X)$
- If $\mu(A,X;\beta)$ is *correct*, $E(Y|A,X)=\mu(A,X;\beta_0)$ for some β_0
- *Estimate* β based on observed data $\Longrightarrow \widehat{\beta}$ (e.g., least squares)
- *Estimate*
  $E\{Y^*(d)\}=E[\mu(1,X,\beta_0)d(X)+\mu(0,X,\beta_0)\{1-d(X)\}]$ by

$$n^{-1}\sum_{i=1}^{n}[\mu(1,X_i,\hat{\beta})d(X_i)+\mu(0,X_i,\hat{\beta})\{1-d(X_i)\}]$$

Estimating the Optimal Regime

- *Estimate* d^{opt} by $\widehat{d}^{opt}_{reg}(X) = I\{\,\mu(1, X; \widehat{\beta}) > \mu(0, X; \widehat{\beta})\,\}$
- "*Regression estimator*"
- Estimator for $E\{Y^*(d^{opt})\}$

$$REG(\widehat{\beta}) = n^{-1} \sum_{i=1}^{n} [\,\mu(1, X_i, \widehat{\beta})\widehat{d}^{opt}_{reg}(X_i) + \mu(0, X_i, \widehat{\beta})\{1 - \widehat{d}^{opt}_{reg}(X_i)\}\,].$$

Concern: $\mu(A, X; \beta)$ may be *misspecified*, so $\widehat{d}^{opt}_{reg}$ could be far from d^{opt}

Estimating the Optimal Regime

Alternative perspective: $\mu(A, X; \beta)$ defines a *class* of regimes

$$d(X, \beta) = I\{\mu(1, X; \beta) > \mu(0, X; \beta)\},$$

indexed by β, that *may or may not* contain d^{opt}

- E.g., suppose *in truth*

$$E(Y|A, X) = \exp\{1 + X_1 + 2X_2 + 3X_1X_2 + A(1 - 2X_1 + X_2)\}$$

$$\Longrightarrow d^{opt}(X) = I(X_2 \geq 2X_1 - 1)$$

Estimating the Optimal Regime

- *Posit*
  $\mu(A, X; \beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)$
- The regimes $I\{\mu(1, X; \beta) > \mu(0, X; \beta)\}$ define a *class* $\mathcal{D}_\eta$ with elements

$$I(X_2 \geq \eta_1 X_1 + \eta_0) \text{ or } I(X_2 \leq \eta_1 X_1 + \eta_0), \quad \eta_0 = -\beta_3/\beta_5,\ \eta_1 = -\beta_4/\beta_5$$

  depending on the sign of β_5
- Notice that the parameter η is defined as a function of β
- The optimal regime in this case is contained in $\mathcal{D}_\eta$
- However, the estimated regime $I\{\mu(1, X; \hat{\beta}) > \mu(0, X; \hat{\beta})\}$ may not estimate the best regime within the class $\mathcal{D}_\eta$ if the posited model is wrong

Optimal Restricted Regime

Suggests: Consider *directly* a restricted set of regimes $\mathcal{D}_\eta = \{d(X, \eta)\}$ *indexed* by η

- Write $d_\eta(X) = d(X, \eta)$
- Such regimes may be motivated by a regression model or based on *cost*, *feasibility* in practice, *interpretability*; e.g., $d(X, \eta) = I(X_1 < \eta_0, X_2 < \eta_1)$
- $\mathcal{D}_\eta$ *may or may not* contain d^{opt} but still of interest
- *Optimal restricted regime* $d_\eta^{opt}(X) = d(X, \eta^{opt})$,

$$\eta^{opt} = \arg\max_\eta E\{Y^*(d_\eta)\}$$

- $\Longrightarrow$ Estimate the optimal restricted regime by *estimating* η^{opt}

Estimating the Optimal Restricted Regime

Approach: Maximize a "*good*" estimator for $E\{Y^*(d_\eta)\}$ in η

- *Missing data* analogy:
- Let C_η denote η-regime consistency indicator; that is,

$$C_\eta = Ad(X, \eta) + (1 - A)\{1 - d(X, \eta)\}$$

- "*Full data*" are $\{Y^*(d_\eta), X\}$; "*observed data*" are $(C_\eta, C_\eta Y, X)$, where
- $\Longrightarrow$ Only a subset of subjects have observed outcomes under $d(X, \eta)$; the rest are *missing*

Estimating the Optimal Restricted Regime

- $\pi(X) = \text{pr}(A = 1|X)$ is the *propensity score* for treatment
- The propensity score is known for randomized studies, or can be estimated using the data $(A_i, X_i), i = 1, \ldots, n$ say using logistic regression $\pi(X; \gamma)$ and estimate γ by $\widehat{\gamma}$.
- The *propensity* of receiving treatment consistent with $d(X, \eta)$

$$\begin{aligned} \pi_c(X; \eta) &= \text{pr}(C_\eta = 1|X) = E(C_\eta|X) \\ &= E[Ad(X, \eta) + (1 - A)\{1 - d(X, \eta)\}|X] \\ &= \pi(X)d(X, \eta) + \{1 - \pi(X)\}\{1 - d(X, \eta)\} \end{aligned}$$

- Write $\pi_c(X; \eta, \gamma)$ with $\pi(X; \gamma)$

Estimating the Optimal Restricted Regime

Estimators for $E\{Y^*(d_\eta)\}$:

- *Inverse probability weighted* estimator

$$IPWE(\eta) = n^{-1} \sum_{i=1}^{n} \frac{C_{\eta,i} Y_i}{\pi_c(X_i; \eta, \widehat{\gamma})}.$$

- *Consistent* for $E\{Y^*(d_\eta)\}$ if $\pi(X;\gamma)$ (hence $\pi_c(X;\eta,\gamma)$) is *correct*

Estimating the Optimal Restricted Regime

- *Doubly robust augmented inverse probability weighted* estimator

$$AIPWE(\eta) = n^{-1} \sum_{i=1}^{n} \left\{ \frac{C_{\eta,i} Y_i}{\pi_c(X_i; \eta, \widehat{\gamma})} - \frac{C_{\eta,i} - \pi_c(X_i; \eta, \widehat{\gamma})}{\pi_c(X_i; \eta, \widehat{\gamma})} m(X_i; \eta, \widehat{\beta}) \right\},$$

$$m(X; \eta, \beta) = E\{Y^*(d_\eta)|X\} = \mu(1, X; \beta) d(X, \eta) + \mu(0, X; \beta)\{1 - d(X, \eta)\}$$

and $\mu(A, X; \beta)$ is a model for $E(Y|A, X)$

- *Consistent* if *either* $\pi(X, \gamma)$ or $\mu(A, X; \beta)$ is *correct*

Augmented Estimator

Under MAR: $Y^*(d_\eta) \perp\!\!\!\perp C_\eta | X$

- If $\widehat{\gamma} \xrightarrow{p} \gamma^*$ and $\widehat{\beta} \xrightarrow{p} \beta^*$, this estimator $\xrightarrow{p}$

$$E\left\{\frac{C_\eta Y}{\pi_c(X;\eta,\gamma^*)} - \frac{C_\eta - \pi_c(X;\eta,\gamma^*)}{\pi_c(X;\eta,\gamma^*)} m(X;\eta,\beta^*)\right\}$$
$$= E\left[Y^*(d_\eta) + \left\{\frac{C_\eta - \pi_c(X;\eta,\gamma^*)}{\pi_c(X;\eta,\gamma^*)}\right\}\{Y^*(d_\eta) - m(X;\eta,\beta^*)\}\right]$$
$$= E\{Y^*(d_\eta)\} + E\left[\left\{\frac{C_\eta - \pi_c(X;\eta,\gamma^*)}{\pi_c(X;\eta,\gamma^*)}\right\}\{Y^*(d_\eta) - m(X;\eta,\beta^*)\}\right]$$

- Hence the estimator is *consistent* if *either*
  - $\pi(X;\gamma^*) = \pi(X) \Rightarrow \pi_c(X;\eta,\gamma^*) = \pi_c(X;\eta)$ (*propensity correct*)
  - $\mu(A,X;\beta^*) = \mu(A,X) \Rightarrow m(X;\eta,\beta^*) = m(X;\eta)$ (*regression correct*)
  - *Double robustness*

Estimating the Optimal Restricted Regime

Result: Estimators $\widehat{\eta}^{opt}$ for η^{opt} obtained by *maximizing* $IPWE(\eta)$ or $AIPWE(\eta)$ in η

- Estimated optimal restricted regime $\widehat{d}_{\eta}^{opt}(X) = d(X, \widehat{\eta}^{opt})$
- *Non-smooth* functions of η; must use suitable *optimization techniques*
- Estimators for $E\{Y^*(d_\eta)\}$

$$IPWE(\widehat{\eta}_{ipwe}^{opt}) \text{ or } AIPWE(\widehat{\eta}_{aipwe}^{opt})$$

  Can calculate *standard errors*
- *Semiparametric theory*: $AIPWE(\eta)$ is *more efficient* than $IPWE(\eta)$ for estimating $E\{Y^*(d_\eta)\}$
- $\Longrightarrow$ Estimating regimes based on $AIPWE(\eta)$ should be "*better*"

Empirical Studies

Extensive simulations: One representative scenario

- *True* $E(Y|A,X)$ of form

$$\mu_t(A,X;\beta) = \exp\{\beta_0+\beta_1 X_1^2+\beta_2 X_2^2+\beta_3 X_1 X_2+A(\beta_4+\beta_5 X_1+\beta_6 X_2)\}$$

- *Misspecified* model for $E(Y|A,X)$

$$\mu_m(A,X;\beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)$$

- $\Longrightarrow \mathcal{D}_\eta = \{I(\eta_0 + \eta_1 X_1 + \eta_2 X_2 > 0)\}$, $d^{opt} \in \mathcal{D}_\eta$
- *True* propensity score $\quad \text{logit}\{\pi_t(X;\gamma)\} = \gamma_0 + \gamma_1 X_1^2 + \gamma_2 X_2^2$
- *Misspecified* propensity score

$$\text{logit}\{\pi_m(X;\gamma)\} = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2$$

Empirical Studies

Both outcome regression models define a class of treatment regimes $\mathcal{D}_\eta = \{I(\eta_0 + \eta_1 X_1 + \eta_2 X_2 > 0)\}$, so that clearly $d^{opt} \in \mathcal{D}_\eta$

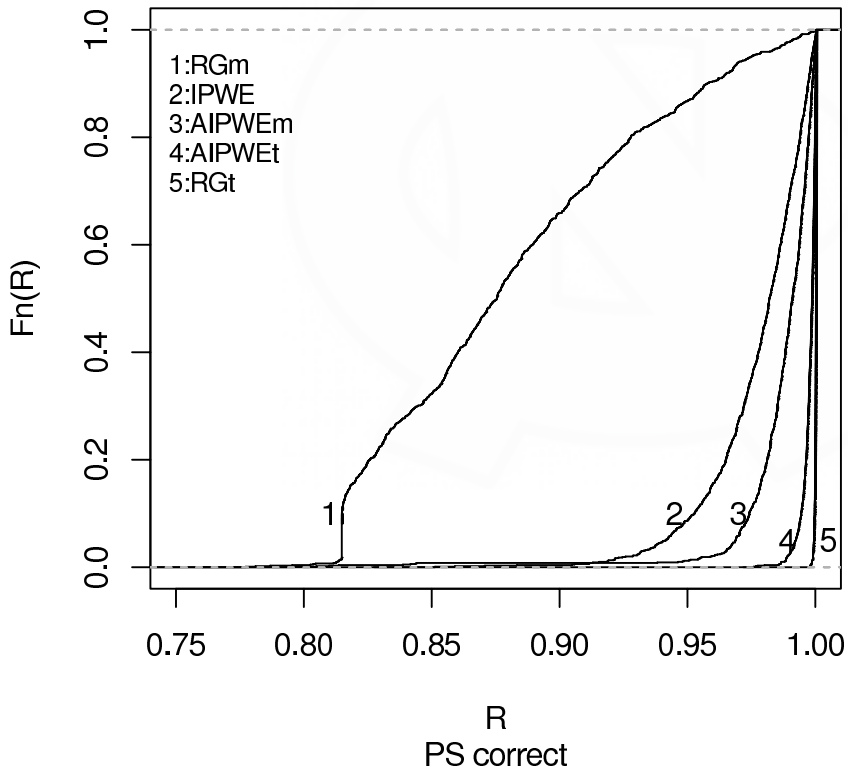
- Expressed in this form, regimes in $\mathcal{D}_\eta$ do not have a unique representation.
- achieved uniqueness by imposing $\| \eta \| = (\eta^T \eta)^{1/2} = 1$.
- In this case, $d^{opt} \in \mathcal{D}_\eta$ corresponds to $\eta = (\eta_0, \eta_1, \eta_2)^T = (-0.07, -0.71, 0.71)^T$

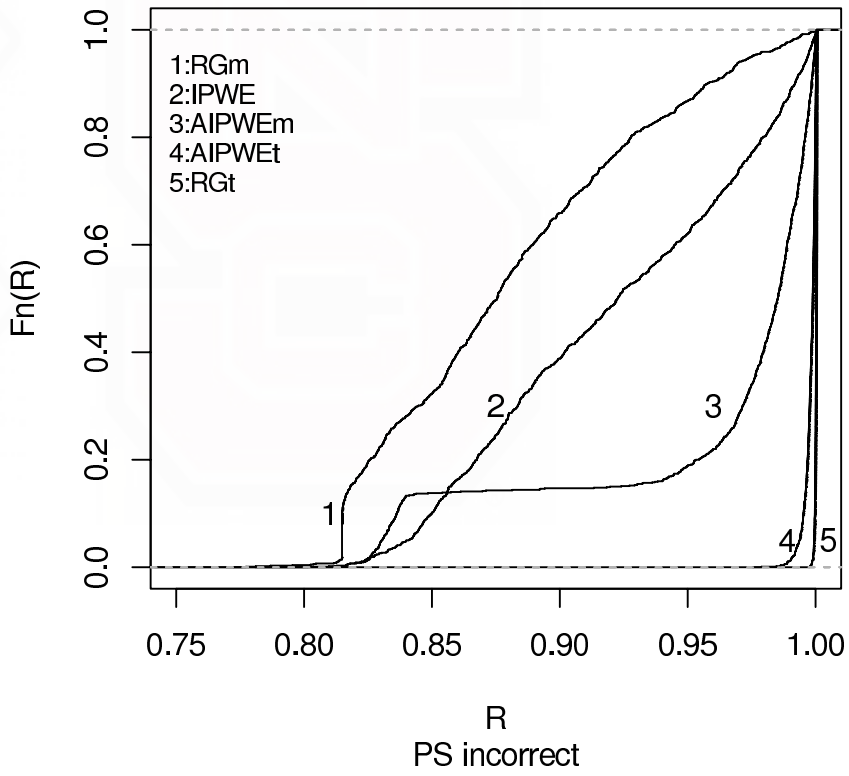
Simulation

- Truth: $(\eta_0, \eta_1, \eta_2) = (-0.07, -0.71, 0.71)$ and $E\{Y^*(d^{opt})\} = 3.71$
- $Q(\eta) = E\{Y^*(d_\eta)\}$, obtained using 10^6 Monte Carlo simulations

Method	$\widehat{\eta}_0$	$\widehat{\eta}_1$	$\widehat{\eta}_2$	$\widehat{E}\{Y^*(d^{opt})\}$	SE	Cov.	$Q(\widehat{\eta}^{opt})$
$RG\mu_t$	-0.07 (0.02)	-0.71 (0.01)	0.71 (0.01)	3.70 (0.14)	–	–	3.71 (0.00)
$RG\mu_m$	-0.51 (0.26)	-0.49 (0.32)	0.46 (0.33)	3.44 (0.18)	–	–	3.27 (0.19)
			PS correct				
$IPWE$	-0.07 (0.15)	-0.69 (0.11)	0.68 (0.11)	4.01 (0.26)	0.28	86.1	3.63 (0.07)
$AIPWE\mu_t$	-0.07 (0.05)	-0.71 (0.03)	0.70 (0.03)	3.72 (0.15)	0.15	94.7	3.70 (0.01)
$AIPWE\mu_m$	-0.06 (0.12)	-0.69 (0.12)	0.69 (0.13)	3.85 (0.21)	0.23	91.8	3.66 (0.07)
			PS incorrect				
$IPWE$	-0.38 (0.22)	-0.56 (0.30)	0.55 (0.31)	4.06 (0.22)	0.23	69.4	3.42 (0.20)
$AIPWE\mu_t$	-0.07 (0.05)	-0.70 (0.02)	0.70 (0.02)	3.72 (0.15)	0.15	95.2	3.70 (0.01)
$AIPWE\mu_m$	-0.23 (0.22)	-0.62 (0.25)	0.61 (0.27)	3.81 (0.18)	0.19	94.1	3.57 (0.20)

Performance: Empirical CDFs of over 1000 data sets of *expected outcome* using $\widehat{d}^{opt}_{reg}$, $\widehat{d}^{opt}(\widehat{\eta}^{opt}_{ipwe})$, $\widehat{d}^{opt}(\widehat{\eta}^{opt}_{aipwe})$ to assign treatment divided by $E(Y^*(d^{opt})\}$ under *true* and *misspecified* models





Application: NSABP Trial

Recall: *Two treatment options*

- $A = 0$ if PFT, $= 1$ if PF
- $Y = 1$ if patient *survived* disease-free to 3 years, $= 0$ otherwise
- $X = (\text{age}, \text{PR})$
- Consider regimes of the form $d(X, \eta) = I(\text{age} < \eta_0 \text{ and } \text{PR} < \eta_1)$
- *Gail and Simon*: $\eta_0 = 50$, $\eta_1 = 10$
- *Estimated optimal regimes*:

	$\widehat{\eta}_0^{opt}$	$\widehat{\eta}_1^{opt}$	Est. $E\{Y^*(d_\eta^{opt})\}$ (95% CI)
IPWE	56	5	0.681 (0.644,0.717)
AIPWE	60	9	0.686 (0.651,0.722)

Discussion

- New methods for estimating an optimal treatment regime within a specified class
- Robustness to misspecification (*AIPWE*)
- Single decision point
- Extension to multiple decisions; is a competitor to Q- and A-learning

References

- Gail, M. and Simon, R. (1985). Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* **41**, 361–372.
- Zhang, B., Tsiatis, A. A., Laber, E.B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, in press.
- Zhang, B. Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, in revision.

Workshop Outline

Introduction to Personalized Medicine and Dynamic Treatment Regimes

Estimation of Optimal Dynamic Treatment Regimes for a Single Decision

Estimation of Optimal Dynamic Treatment Regimes for Multiple Decisions

Advanced Topics in Personalized Medicine and Dynamic Treatment Regimes

Dynamic Treatment Regimes

- Motivation : treatment of chronic illness
  - Some examples: HIV/AIDS, cancer, depression, schizophrenia, drug and alcohol addiction, ADHD, etc.
  - Multistage decision making problem
  - Longer-term treatment requires consideration and tradeoff of present versus longer term benefit.
- Dynamic treatment regimes (DTRs)
  - Operationalize multistage decision making via as sequence of decision rules
    - One decision rule for each time (decision) point
    - A decision rule is a function inputs patient history and outputs a recommended treatment
  - Aim to optimize some cumulative clinical outcome

Related Problems

- Construction and inference for policies have applications beyond medicine
  1. Artificial Intelligence and Reinforcement Learning (autonomous helicopter, drones, etc., Ng 2003)
  2. Marketing (Simester, Sun and Tsitsiklis, 2003)
  3. Active labor market policies (Lechner and Miquel, 2010)
  4. Adaptive learning for games (tux cart, plants vs. zombies)
  5. ...

Roadmap

1. Two examples of SMARTs
2. Q-learning
3. Whirlwind tour of known issues

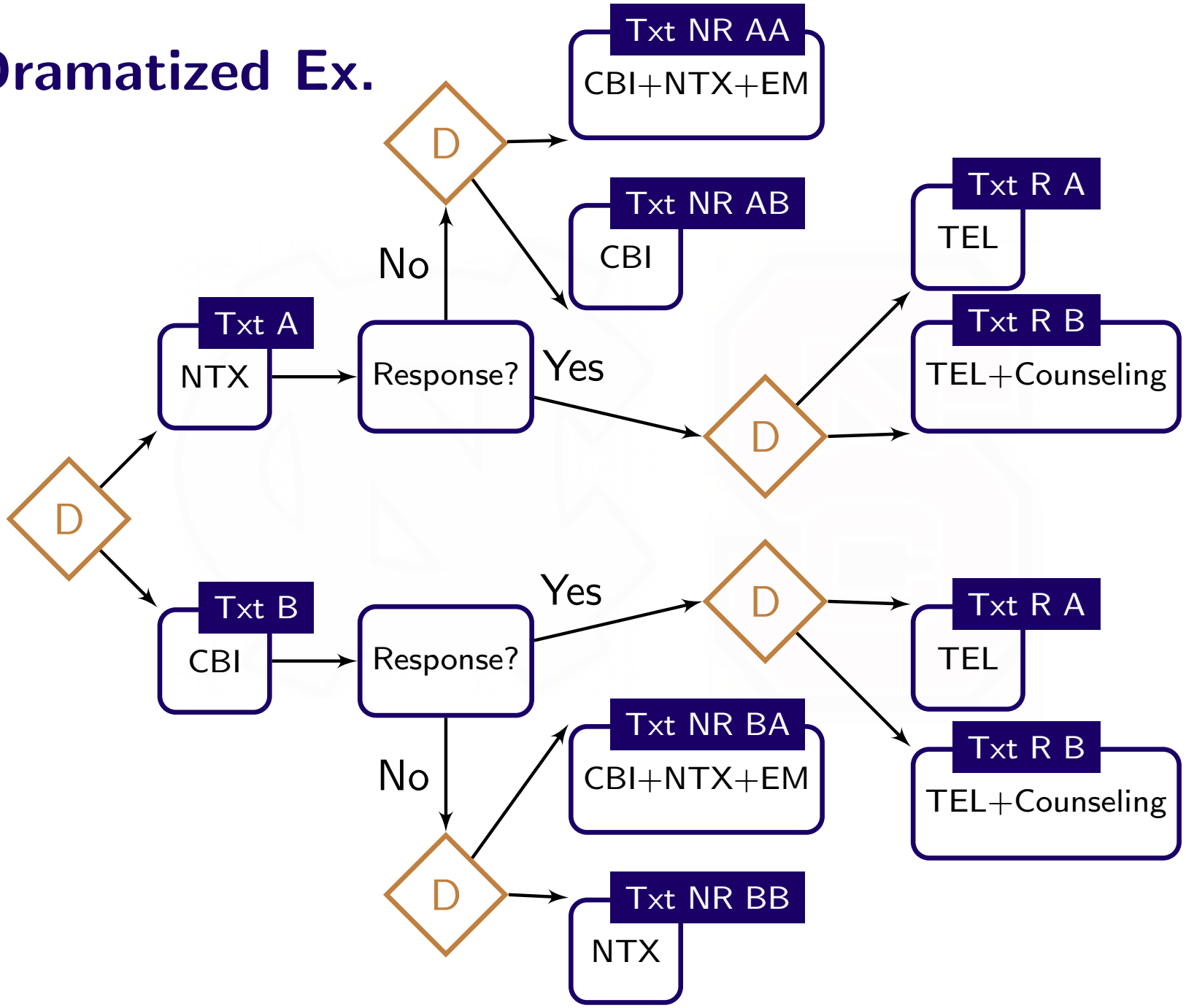
Roadmap

1. Two examples of SMARTs
2. Q-learning
3. Whirlwind tour of known issues

Dramatized Example

- Addiction management example inspired by the ExTENd and COMBINE trials (Murphy, 2005, Qian et. al., 2012)
- Devising two-time point txt strategy for alcohol dependent patients
  - Initial txt choices Naltrexone (NTX) and Combined Behavioral Intervention (CBI)
  - At six-months responders classified as responders or non-responders
  - For responders to initial txt, followup txt choices are telephone monitoring (TEL) and telephone monitoring + counseling (TEL+Counseling)
  - For non-responders to initial txt, followup txt choices are switch initial txts (NTX $\leftrightarrow$ CBT), or step-up initial txt CBI + NTX + Enhanced monitoring (CBI + NTX +EM)
- Primary outcome: percent days abstinent in one year

Dramatized Ex.



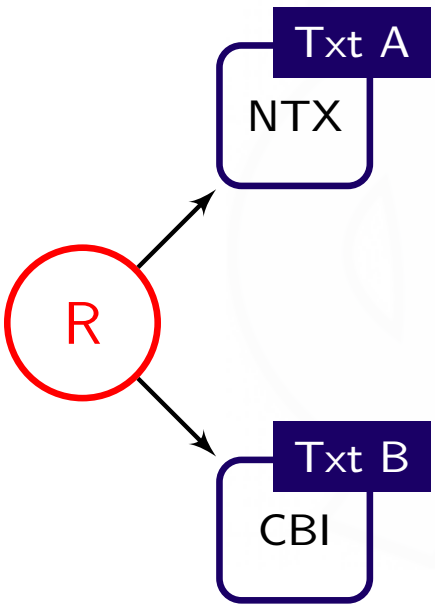
Examples of Simple Treatment Regimes

- Regime 1: Prescribe NTX initially; then assign TEL to responders; and assign step-up to non-responders.
- Regime 2: Prescribe CBI initially; then assign TEL+Counseling to responders; and assign step-up to non-responders.

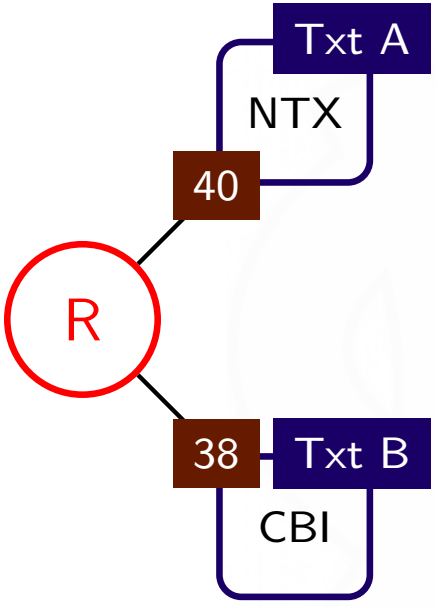
Choosing a Regime

- If we do not take into account individual patient characteristics, there are 8 possible regimes. How can we empirically estimate the best treatment regime?
- Myopic approach
  1. Conduct two-arm trial of NTX vs CBI, pick 'winner' based on mean comparison
  2. Conduct a follow-up study that initially assigns the 'winner' from step 1, then randomizes responders to either TEL or TEL + Counseling, and randomizes non-responders to step-up or switch. Choose 'winners' within the responder and non-responder groups using mean comparison.

Myopic Approach: Step 1

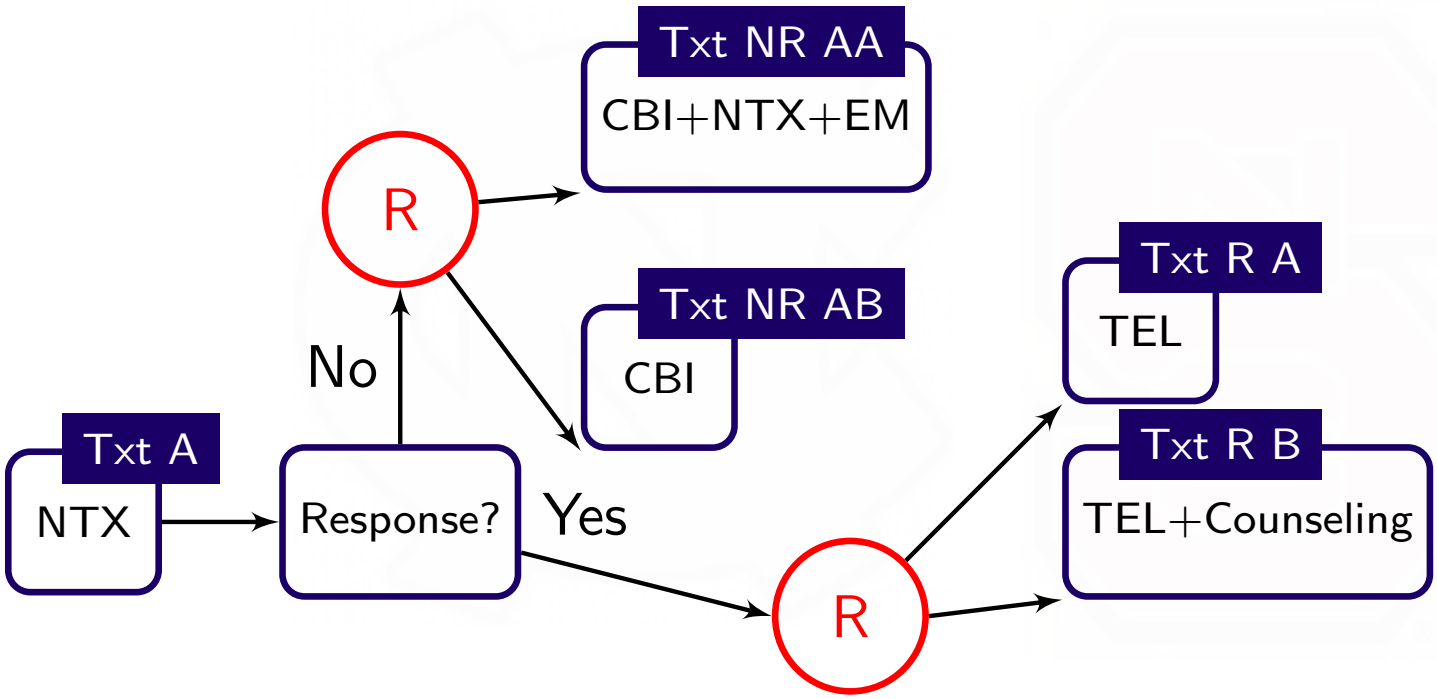


Myopic Approach: Step 1

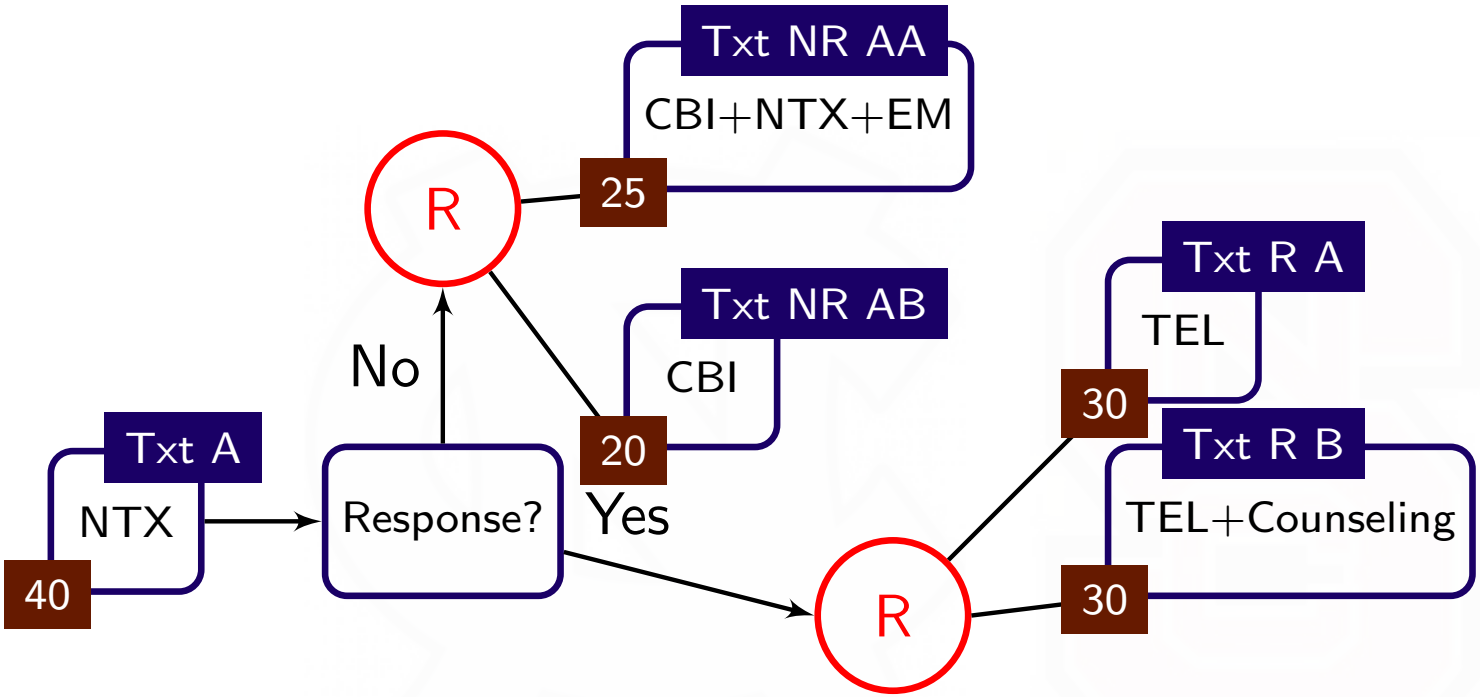


- Nbrs in red denote days abstinent in six-month period
- NTX is yields better immediate six-month outcome

Myopic Approach: Step 2

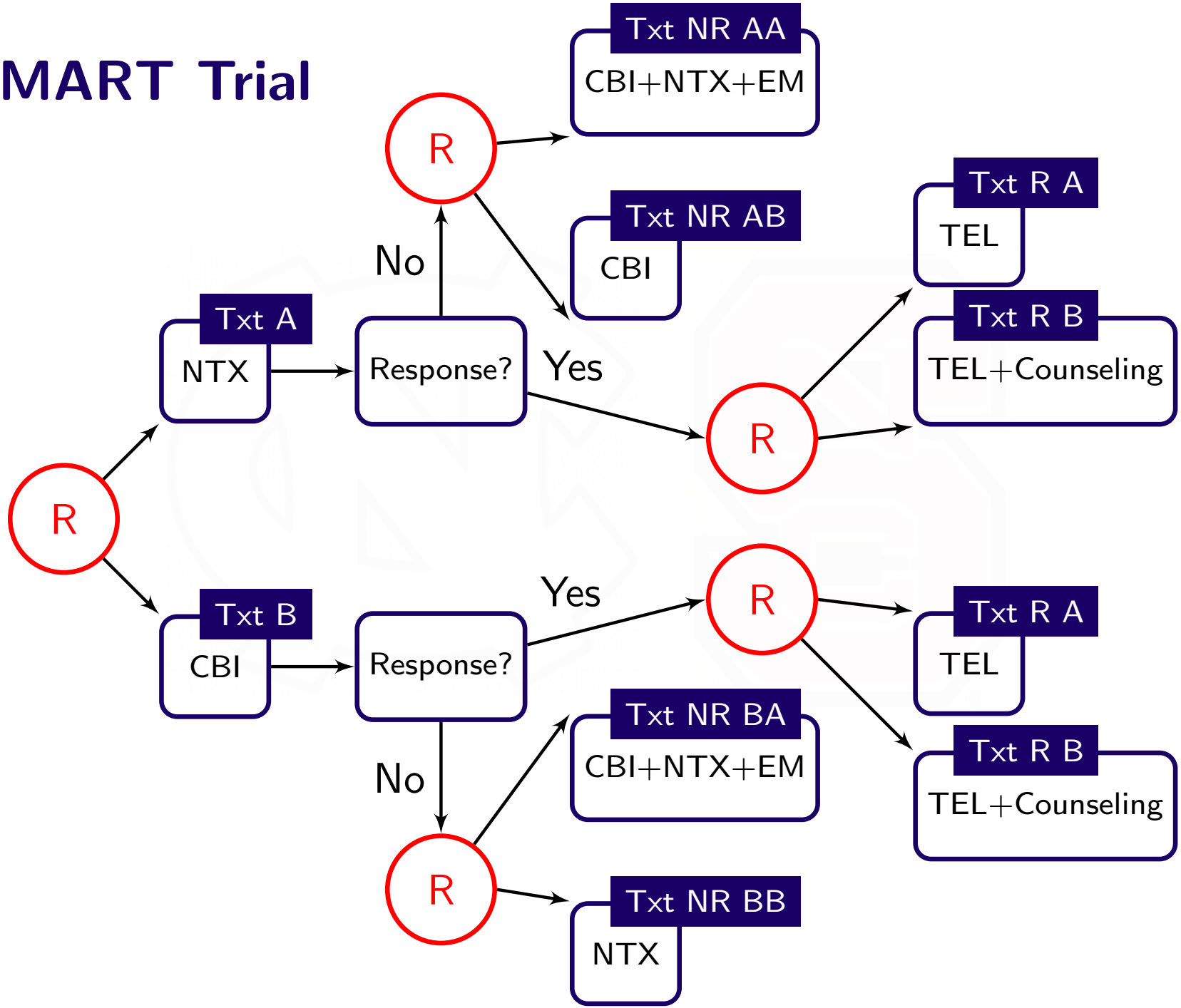


Myopic Approach: Step 2

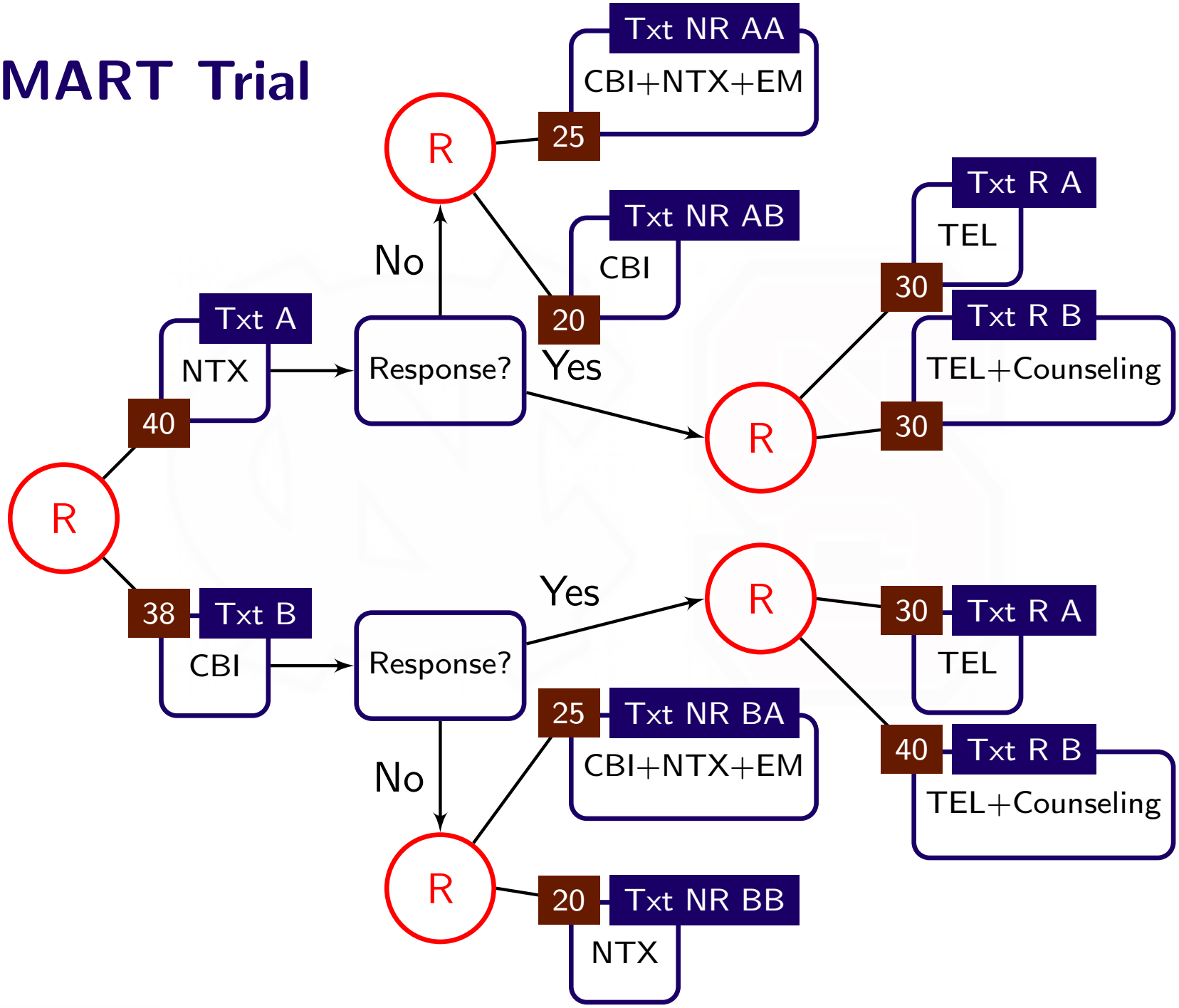


- Myopic regime: Initially prescribe NTX, then assign step-up for non-responders, and assign TEL for responders
  - Assuming that 50% of patients respond, this regime results in an average of 33.75% days abstinent over a one-year period
- Is this optimal among the eight regimes considered?

SMART Trial


R
Txt A
NTX
Response?
No
R
Txt NR AA
CBI+NTX+EM
Txt NR AB
CBI
Yes
R
Txt R A
TEL
Txt R B
TEL+Counseling
R
Txt B
CBI
Response?
Yes
R
Txt R A
TEL
Txt R B
TEL+Counseling
No
R
Txt NR BA
CBI+NTX+EM
Txt NR BB
NTX

SMART Trial


R
25
Txt NR AA
CBI+NTX+EM
No
20
Txt NR AB
CBI
Txt A
NTX
Response?
Yes
40
R
30
Txt R A
TEL
Txt R B
TEL+Counseling
30
R
38
Txt B
CBI
Response?
Yes
R
30
Txt R A
TEL
No
25
Txt NR BA
CBI+NTX+EM
40
Txt R B
TEL+Counseling
R
20
Txt NR BB
NTX

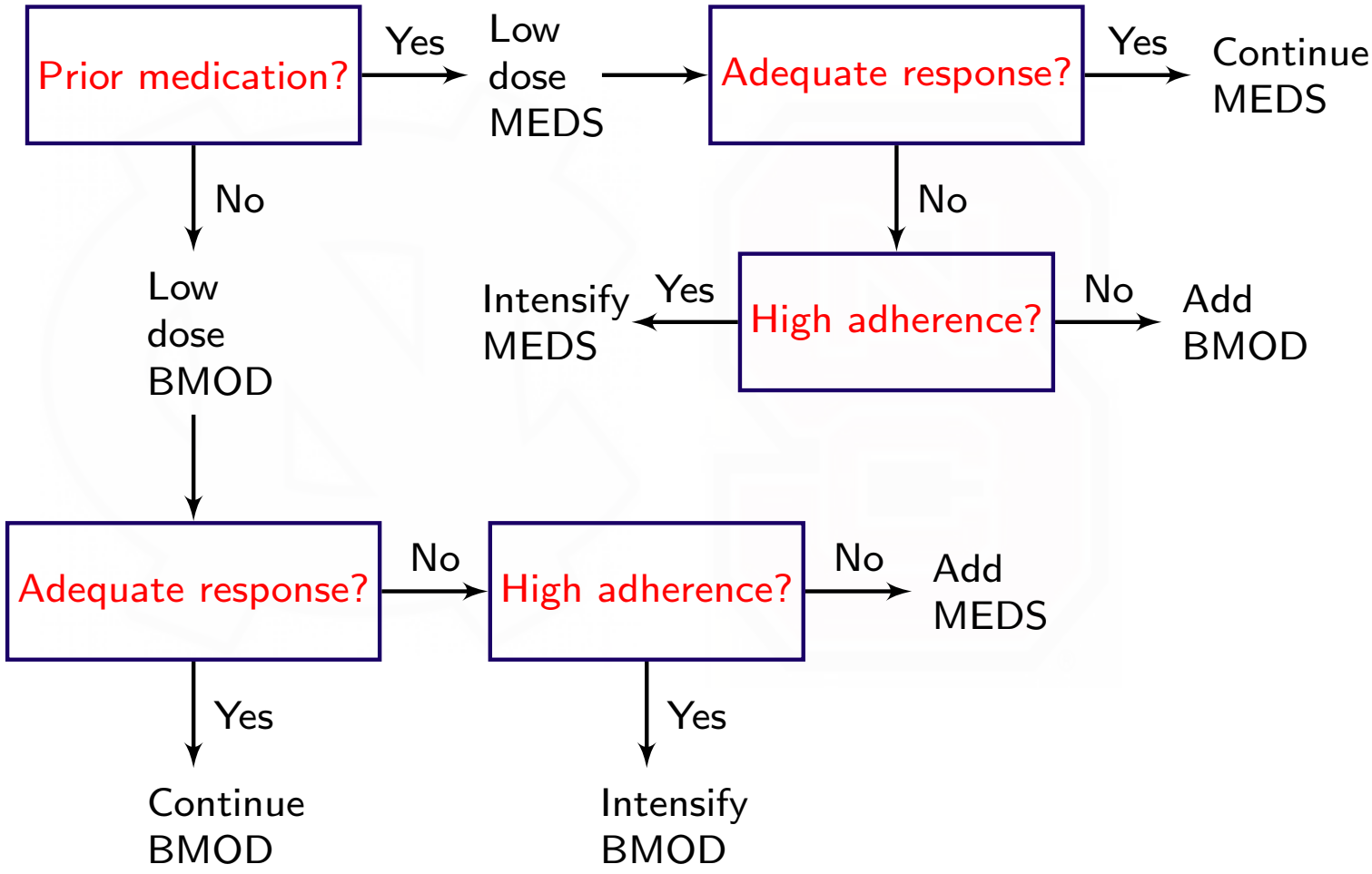
Delayed Effects

- Optimal regime: initially assign CBI, then assign TEL+Counseling to responders, and step-up to non-responders
  - Assuming 50% of patients respond, this regime results in an average of 35.25% days abstinent over a one-year period.
- Myopic regime results in suboptimal patient care
  - Giving CBI initially taught responders to more effectively use counseling yielding better long term outcomes
  - This is delayed effect of assigning CBI

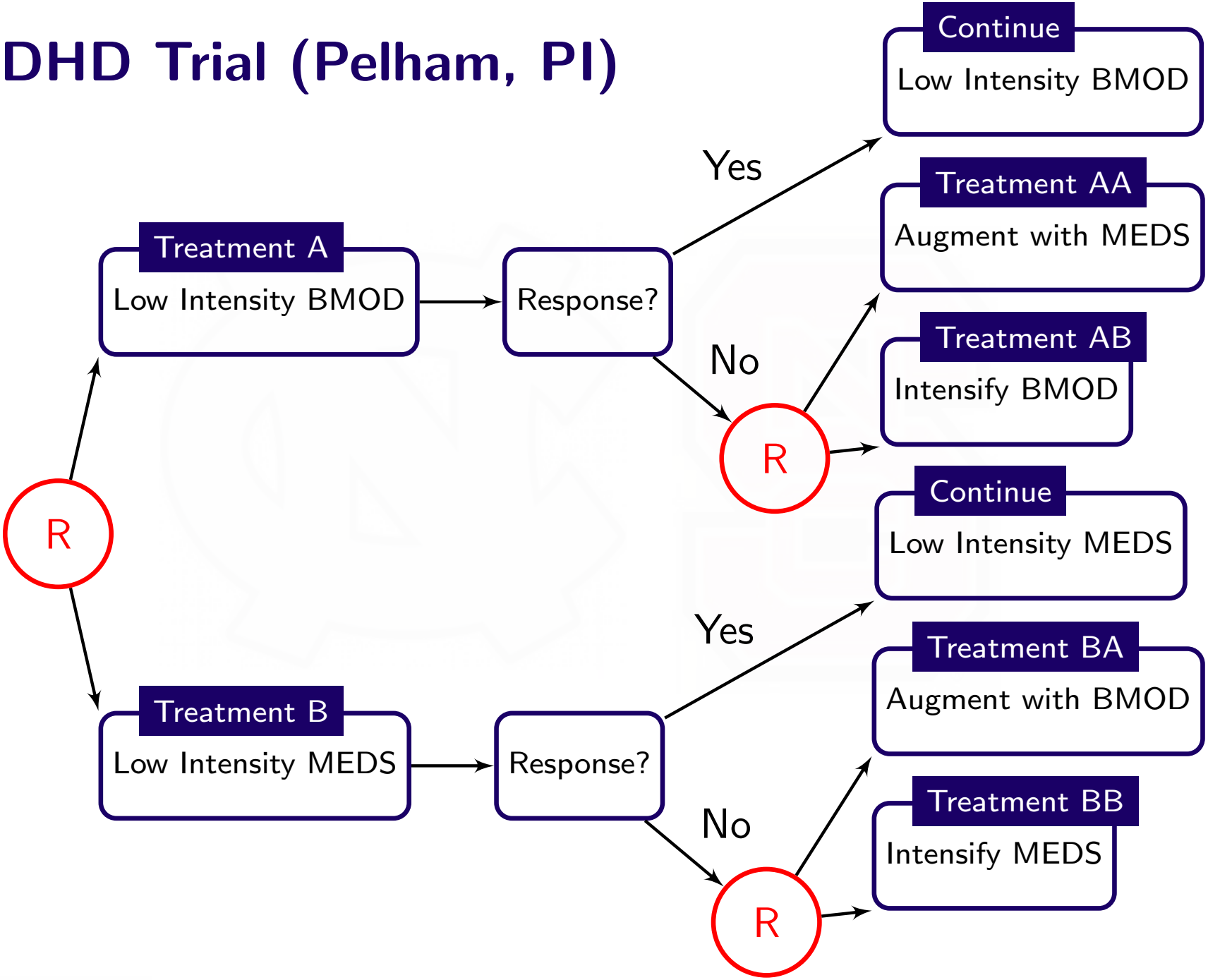
Delayed Effects, cont'd

- Chronic illness requires consideration of treatment sequences
- Must accommodate intermediate information including prior txts into current txt choice
  - Delayed effects
  - Berkson's fallacy (see Gail and Benichou, 2000)

An Example Policy for ADHD



ADHD Trial (Pelham, PI)



Roadmap

1. Two examples of SMARTs
2. Q-learning
3. Whirlwind tour of known issues

Data

- (X_1, A_1, X_2, A_2, Y) for each individual
  - X_k: Observations available at stage k
  - A_k: Treatment at stage k
  - Y: Primary outcome (larger is better)
  - H_k: History at stage k, $H_1 = X_1$, $H_2 = (X_1, A_1, X_2)$
- The regime, $d = \{d_1, d_2\}$, $d_k : \mathcal{H}_k \to \mathcal{A}_k$, should have high Value: $V^d = E^d(Y)$
  - The value corresponds to the average outcome if all patients are assigned treatment according to d
  - Optimal decision rule d^{opt} satisfies $\mathbb{E}^{d^{\text{opt}}} Y = \sup_d \mathbb{E}^d Y$

Review: Dynamic Programming

- Optimal regime d^{opt} can be derived using dynamic programming (Bellman, 1957)
  - Define
    - $Q_2(h_2, a_2) \triangleq \mathbb{E}\left(Y \middle| H_2 = h_2, A_2 = a_2\right)$
    - $Y^* \triangleq \max_{a_2} Q_2(H_2, a_2)$
    - $Q_1(h_1, a_1) \triangleq \mathbb{E}\left(Y^* \middle| H_1 = h_1, A_1 = a_1\right)$
  - $d_k^{\text{opt}}(h_K) = \arg\max_{a_k} Q_k(h_k, a_k)$

Constructing a DTR from Data: Q-Learning

- When system dynamics are known dynamic programming yields the optimal DTR
- We only have data!
- Q-learning mimics dynamic programming but replaces conditional expectations with (typically linear) regression models

Simple Version of Q-Learning

Two stages; linear regressions; $A_k \in \{0, 1\}$, H_{k1}, H_{k2} features of patient history, H_k:

- Stage 2 regression: Regress Y on H_{21}, H_{22} to obtain $\hat{Q}_2(H_2, A_2) = \hat{\beta}_{21}^T H_{21} + \hat{\beta}_{22}^T H_{22} A_2$
  - $\hat{d}_2(H_2) = \arg\max_{a_2} \hat{Q}_2(H_2, a_2) = \arg\max_{a_2} \hat{\beta}_{22}^T H_{22} a_2$

Simple Version of Q-Learning

Two stages; linear regressions; $A_k \in \{0, 1\}$, H_{k1}, H_{k2} features of patient history, H_k:

- Stage 2 regression: Regress Y on H_{21}, H_{22} to obtain $\hat{Q}_2(H_2, A_2) = \hat{\beta}_{21}^T H_{21} + \hat{\beta}_{22}^T H_{22} A_2$
  - $\hat{d}_2(H_2) = \arg\max_{a_2} \hat{Q}_2(H_2, a_2) = \arg\max_{a_2} \hat{\beta}_{22}^T H_{22} a_2$
- $\tilde{Y} = \hat{\beta}_{21}^T H_{21} + \max_{a_2} \hat{\beta}_{22}^T H_{22} a_2$
  - $\tilde{Y}$ is a predictor of $\max_{a_2} Q_2(H_2, a_2)$

Simple Version of Q-Learning

Two stages; linear regressions; $A_k \in \{0, 1\}$, H_{k1}, H_{k2} features of patient history, H_k:

- Stage 2 regression: Regress Y on H_{21}, H_{22} to obtain $\hat{Q}_2(H_2, A_2) = \hat{\beta}_{21}^T H_{21} + \hat{\beta}_{22}^T H_{22} A_2$
  - $\hat{d}_2(H_2) = \arg\max_{a_2} \hat{Q}_2(H_2, a_2) = \arg\max_{a_2} \hat{\beta}_{22}^T H_{22} a_2$
- $\tilde{Y} = \hat{\beta}_{21}^T H_{21} + \max_{a_2} \hat{\beta}_{22}^T H_{22} a_2$
  - $\tilde{Y}$ is a predictor of $\max_{a_2} Q_2(H_2, a_2)$
- Stage 1 regression: Regress $\tilde{Y}$ on H_{11}, H_{12} to obtain $\hat{Q}_1(H_1, A_1) = \hat{\beta}_{11}^T H_{11} + \hat{\beta}_{12}^T H_{12} A_1$
  - $\hat{d}_1(H_{12}) = \arg\max_{a_1} \hat{Q}_1(H_1, a_1) = \arg\max_{a_1} \hat{\beta}_{12}^T H_{12} a_1$

Q-learning Positives

- Natural approximate dynamic programming approach
- Linear models are common but non-essential
  - Parsimonious and interpretable
  - More flexible models can be used to define the Q-functions (e.g., boosting, random forests, etc.)
- Regression models are well-understood
  - Diagnostic and validation tools exist
  - EDA is straightforward

Roadmap

1. Two examples of SMARTs
2. Q-learning
3. Whirlwind tour of known issues

Q-learning ... Opportunities

- Non-smooth non-monotone max-operator
  - Linear models are rarely correctly specified for Q_1
  - Non-smoothness induces non-regularity so that standard methods for inference, e.g., the bootstrap and taylor series arguments, are invalid
  - Non-monotone transformations are difficult to model
- Q-learning indirectly estimates d^{opt} through the conditional mean functions
  - Recall, $d_k^{\text{opt}} = \arg\max_{a_k} Q_k(h_k, a_k)$ which depends only on the sign of $Q_k(h_k, 1) - Q_k(h_k, 0)$.
  - Analog in classification: logistic classification vs. large-margin classification

Linear Models are Rarely Correctly Specified for Q_1

- Toy generative model

$$\begin{aligned} X_1 &\sim \text{Normal}(0,1), & \xi &\sim \text{Normal}(0,1/2), \\ X_2 &= \zeta X_1 + \xi, & A_K &\sim \text{Uniform}\{0,1\}, k = 1,2, \\ \phi &\sim \text{Normal}(0,1/2), & Y &= 1.25 A_1 A_2 + A_2 X_2 - A_1 X_1 + \phi, \end{aligned}$$

  ζ governs the correlation between X_1 and X_2

- Linear model is correct for Q_2

$$Q_2(H_2, A_2) = 1.25 A_1 A_2 + A_2 X_2 - A_2 X_1$$

- Nonlinear model required for Q_1

$$\begin{aligned} Q_1(H_1, A_1) = \frac{1}{2\sqrt{2\pi}} \exp\left\{-2(1.25A_1 + \zeta X_1)^2\right\} \\ + (1.25A_1 + \zeta X_1)\Phi\left(2(1.25A_1 + \zeta X_1)\right) \end{aligned}$$

Linear Models are Rrarely . . . cont'd

- Nonlinear model required for Q_1

$$Q_1(H_1, A_1) = \frac{1}{2\sqrt{2\pi}} \exp\left\{-2(1.25A_1 + \zeta X_1)^2\right\} + (1.25A_1 + \zeta X_1)\Phi\left(2(1.25A_1 + \zeta X_1)\right)$$

- This is an idealized setting, yet:
  - Linear model assumption holds only when $\zeta = 0$, but this is unlikely in practice
  - Even seasoned data analysts would likely have trouble identifying the correct functional form given limited data

Non-smoothness Invalidates Std Inference

- Due to max-operator, $Q_1(h_1, a_1)$ is a non-smooth functional of the generative distribution
  - No regular estimators of Q_1 exist
  - No asymptotically unbiased estimators of $Q_1(h_1, a_1)$ exist
  - Estimators do not converge uniformly over parameter space; standard approaches like bootstrap and Taylor series arguments are invalid
- There is now a small industry built around trying to alleviate some of these problems
  - Thresholding and penalized methods (Moodie and Richardson, 2009; Chakraborty et al., 2010; Song et al., 2012)
  - Local asymptotic approaches (Laber et al., 2012; `SAS` PROC QLEARN)
  - Resampling approaches (Chakraborty et al., 2012; `R` package qLearn)

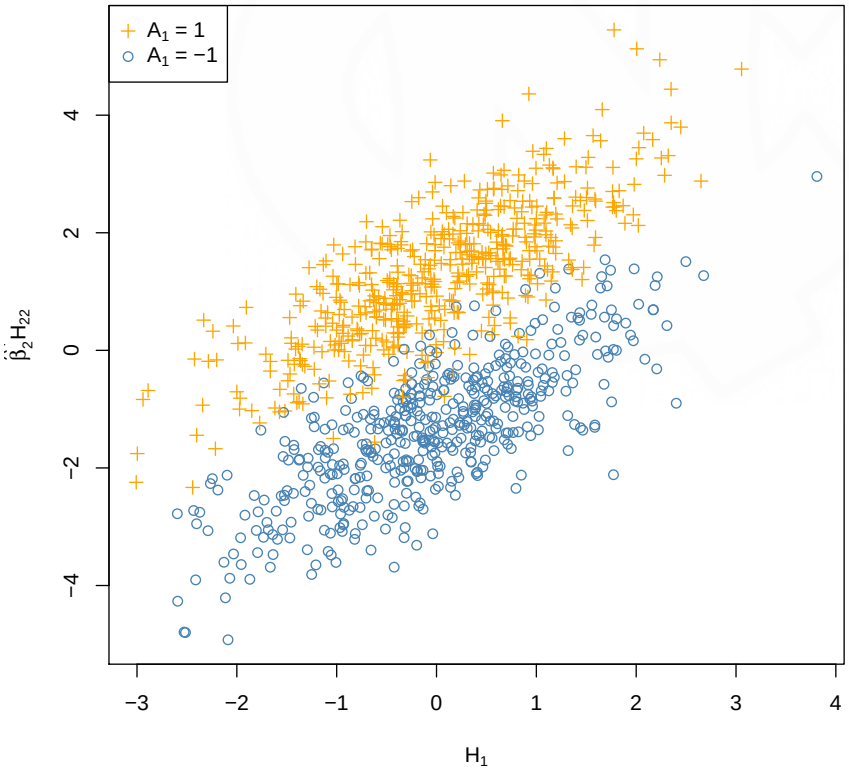
Non-smooth Mon-monotone Transformations

- Recall $\tilde{Y} = \max_{a_2} \hat{Q}_2(H_2, a_2) = \hat{\beta}_{21}^{\mathsf{T}} H_{21} + \max(\hat{\beta}_{22}^{\mathsf{T}} H_{22}, 0)$

Non-smooth Mon-monotone Transformations

- Recall $\tilde{Y} = \max_{a_2} \hat{Q}_2(H_2, a_2) = \hat{\beta}_{21}^{\mathsf{T}} H_{21} + \max(\hat{\beta}_{22}^{\mathsf{T}} H_{22}, 0)$

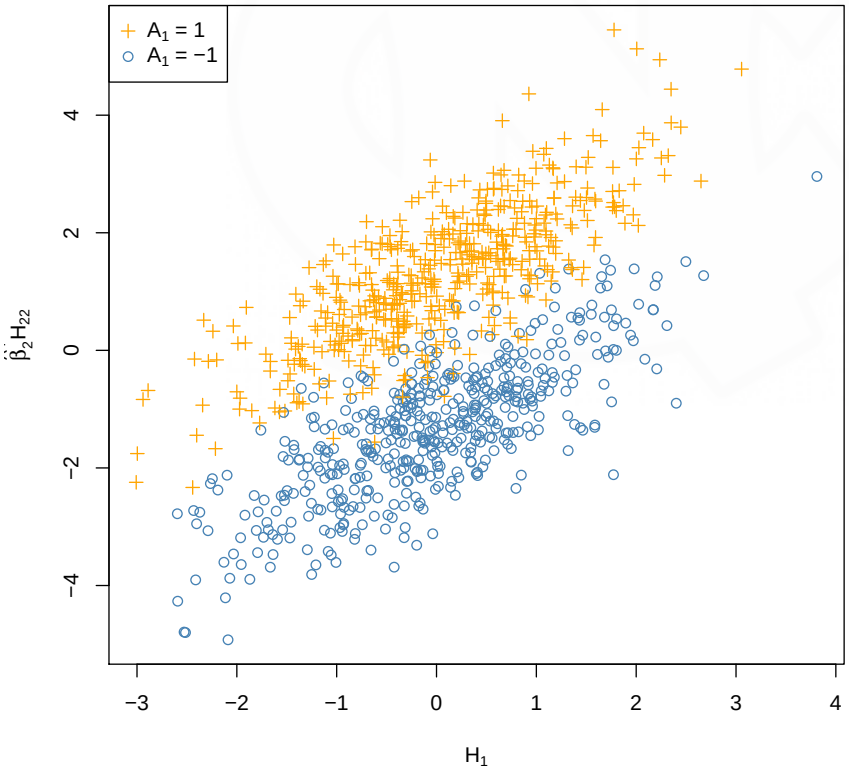
Before maximization



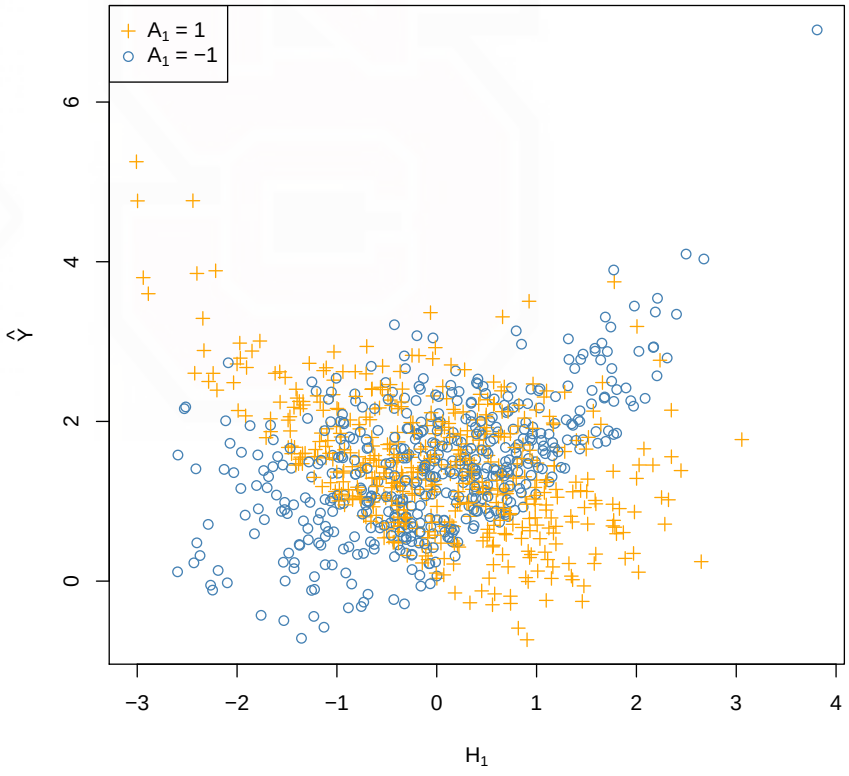
Non-smooth Mon-monotone Transformations

- Recall $\tilde{Y} = \max_{a_2} \hat{Q}_2(H_2, a_2) = \hat{\beta}_{21}^{\mathsf{T}} H_{21} + \max(\hat{\beta}_{22}^{\mathsf{T}} H_{22}, 0)$

Before maximization



After maximization



Non-smooth Non-monotone Transformations, cont'd

- Dealing with non-smooth, non-monotone transformations is difficult in practice
- On approach is to interchange modeling and maximization
  - Only need to model smooth transformations of the data
  - Requires more modeling (see Laber et al., 2012)

Q-learning Indirectly Estimates d^{opt}

- $d_k^{\text{opt}}(h_k) = \arg\max_{a_k} Q_k(h_k, a_k) = \mathbf{1}_{Q_k(h_k,1)-Q_k(h_k,0)>0}$
- Thus, $d_k^{\text{opt}}(h_k)$ depends only on the sign of contrast $Q_k(h_k, 1) - Q_k(h_k, 0)$
  - Q-learning estimates $Q_k(h_k, a_K)$, hence does not directly target d^{opt}
  - A-learning (Murphy, 2003; Murphy et al., 2004) targets $Q_k(h_k, 1) - Q_k(h_k, 0)$, is closer but still indirect
- Recent classification-based estimators of Zhang et al. (2012), Zhao et al. (2012), and Xi et al. (2012) directly target d^{opt}
  - These are multistage extensions of what Butch introduced in the previous segment

Classification Estimators

- For clarity, simplify development of Zhao et al. (2012)
  - Assume Y is nonnegative
  - Assume A_1 and A_2 are randomly assigned as in a SMART
  - Recode A_k to take values in $\{-1, 1\}$
- For any policy d the value equals

$$\mathbb{E}^d Y = \mathbb{E}\left(\frac{Y\mathbf{1}_{A_2=d_2(H_2)}\mathbf{1}_{A_1=d_1(H_1)}}{p(A_1|H_1)p(A_2|H_2)}\right)$$

- Empirical analog

$$\frac{1}{n}\sum_{i=1}^{n}\omega_i\mathbf{1}_{\min\{A_{2i}d_2(H_{2i}),\, A_{1i}d_1(H_{1i})\}\geq 0},$$

where $w_i \triangleq Y_i/p(A_{1i}|H_{1i})p(A_{2i}|H_{2i})$

Classification Estimators, cont'd

- Empirical analog

$$\frac{1}{n}\sum_{i=1}^{n}\omega_i \mathbf{1}_{\min\{A_{2i}d_2(H_{2i}),\, A_{1i}d_1(H_{1i})\}\geq 0},$$

  where $w_i \triangleq Y_i / p(A_{1i}|H_{1i})p(A_{2i}|H_{2i})$
- Similar weighted misclassification rate
  - New 'margin' $\min(A_2 d_2(H_2), A_1 d_1(H_1))$
  - Weights $\omega = Y / p(A_1|H_1)p(A_2|H_2)$
- Classification estimators (approximately) maximize the empirical value over $d = (d_1, d_2)$ in $\mathcal{D}$
  - Zhao et al. (2012) employ SVMs
  - Zhang et al. (2012) use a genetic algorithm to maximize an augment version of the empirical value
  - Xi et al. (2012) use convex surrogates and an augmented version of the empirical value

Classification Estimators, cont'd

- Classification estimators directly target the decision rule
- Loss of prognostic information
- Directly minimizing the empirical value is computationally difficult
- Replacing indicator with a convex surrogate may lead to suboptimal solutions unless model space is correct

Wrap-up

- This is an extremely active area of research
- Tools for estimation and inference exist and are continually being improved
- There is no panacea, choosing the proper statistical tool depends critically on the goals of the analysis
- There is help!
  - Statisticians on the P01
  - UNC-NCSU working group on dynamic treatment regimes
  - NCSU personalized medicine cluster

Workshop Outline

Introduction to Personalized Medicine and Dynamic Treatment Regimes

Estimation of Optimal Dynamic Treatment Regimes for a Single Decision

Estimation of Optimal Dynamic Treatment Regimes for Multiple Decisions

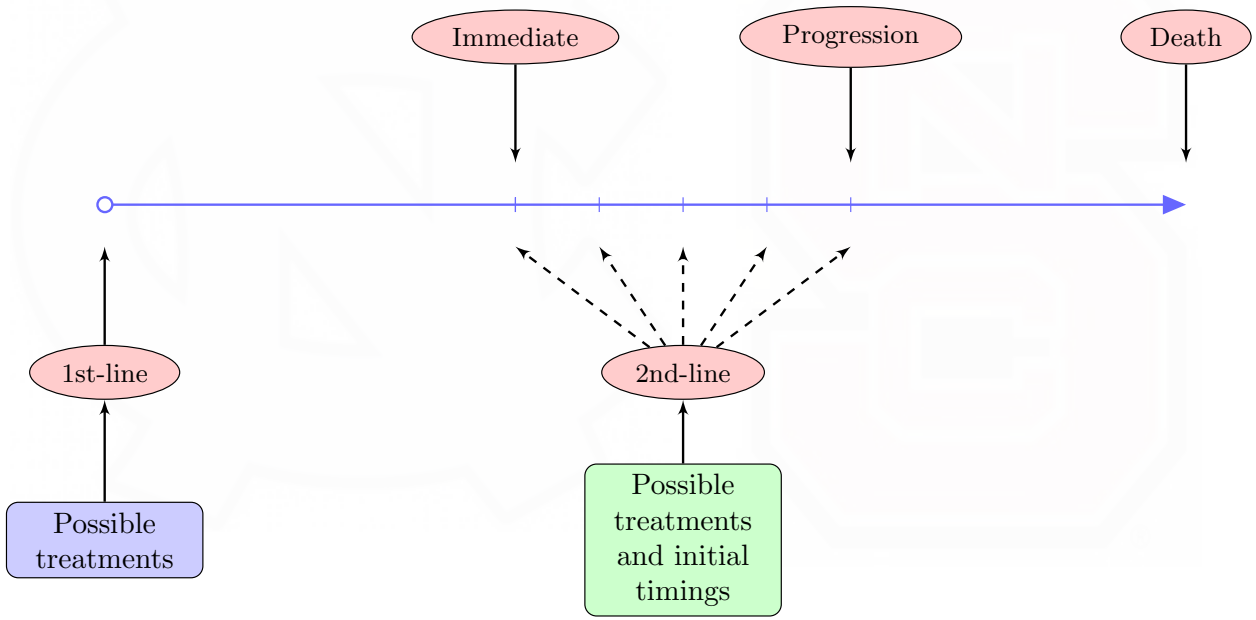
Advanced Topics in Personalized Medicine and Dynamic Treatment Regimes

Outline

- Some Illustrative Examples
- Overview of Statistical Issues
- Statistical Learning
- Incorporating Censoring
- Outcome Weighted Learning
- Open Questions
- Preparing Protocols

Example 1: Non-Small Cell Lung Cancer

In treating advanced non-small cell lung cancer, patients typically experience two or more lines of treatment.



Problem of Interest

Can we improve survival by personalizing the treatment at each decision point (drug at both and timing at second) based on prognostic data?

Example 1: Non-Small Cell Lung Cancer

The clinical setting:

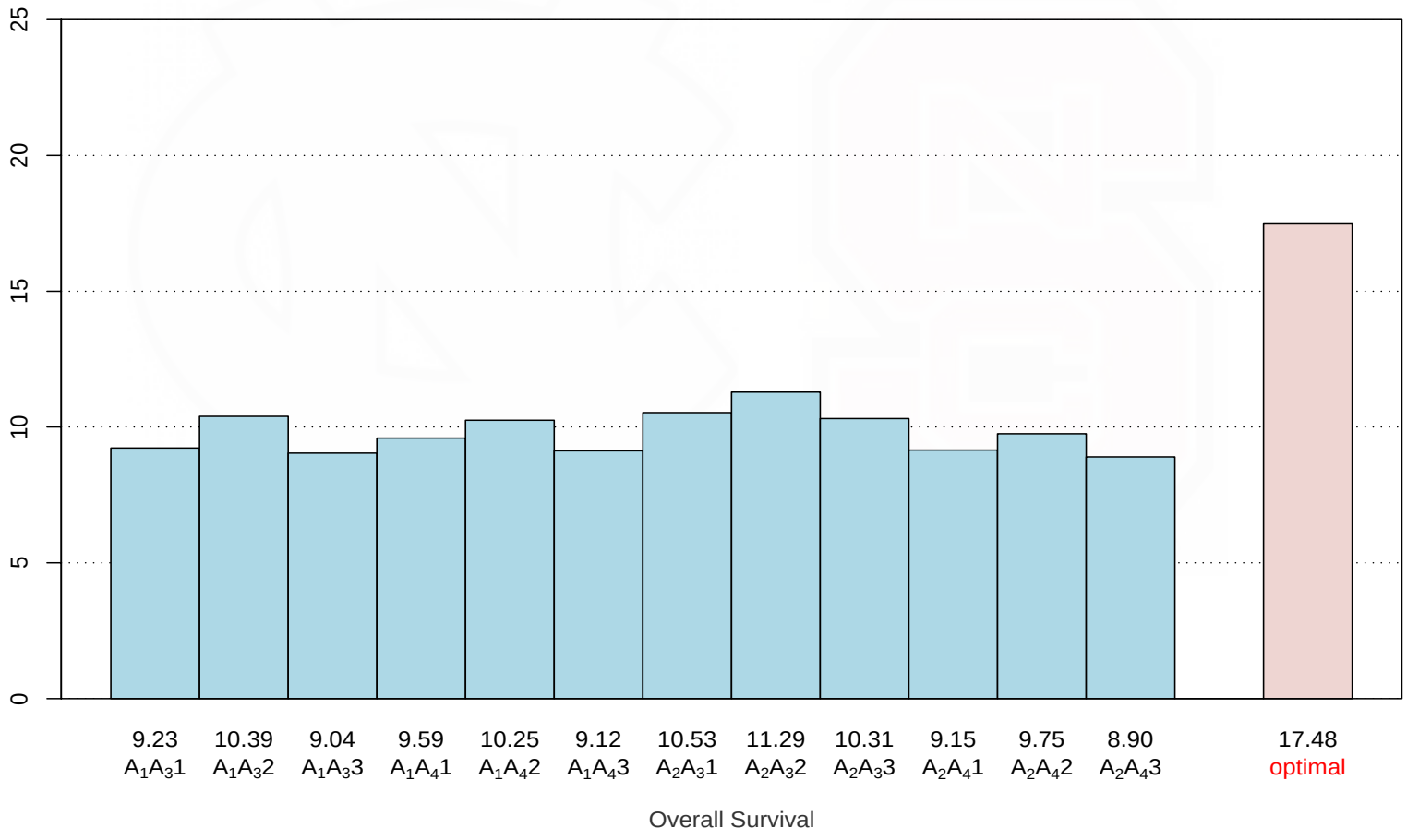
- There are two to three lines of therapy, but very few utilize three, and we will focus on two here.
- We need to make decisions at two treatment times: (1) at the beginning of the first line and (2) at the end of the first line.
- For time (1), we need to decide which of several agent options is best: we will only consider two options in the simulation.
- For time (2), we need to decide when to start the second line (out of three choices for simplicity) and which of two agents to assign.
- The reward function is overall survival which is right-censored.

Example 1: Non-Small Cell Lung Cancer

Realistic simulated patients (Zhao, et al., 2011):

- Difference equations used to generate patient trajectories for two clinical measures: tumor size and quality of life.
- Four distinct subgroups formed with different relationships between treatment (agents and timing) and measures.
- A SMART trial was simulated using two different drugs at each decision point and three different timings at the second point, yielding 12 different treatment pathways.
- Q-learning was used to estimate decision rules based on treatment history and clinical measures as tailoring variables and survival time as clinical outcome.
- A Phase III trial comparing the 12 treatment paths with the estimated optimal individualized decision rules was simulated.

Example 1: Non-Small Cell Lung Cancer



Example 1: Non-Small Cell Lung Cancer

Some statistical issues:

- Statistical learning is very useful for handling
  - nonlinear structure,
  - complex interactions, and
  - large numbers of variables.
- Statistical learning tools for censored data are very limited (almost nonexistent) and appropriate extensions are needed.
- Complex treatment decisions (involving multiple drugs and/or timing) are new challenges for statistical learning.

Example 2: Bronchopulmonary Dysplasia in Infants

The clinical setting:

- Sildenafil has been shown to be effective in preventing bronchopulmonary dysplasia-associated pulmonary hypertension in premature infants.
- A crucial open question is what dose to use with which patients.
- We designed a Phase II dose finding study with the intent of achieving individualized dosing rules.

Example 2: Bronchopulmonary Dysplasia in Infants

Scientific and statistical issues:

- The investigators would like the design to be adaptive so that ineffective or harmful doses are discarded early.
- A challenge for statistical learning is that dose is a continuous treatment decision.
- Since the methodology is new and unfamiliar, how do we frame the design and proposal in a manner that it will satisfy reviewers and obtain approval?

Example 3: Cystic Fibrosis

The clinical setting:

- Cystic fibrosis (CF) is a genetic disease.
- The most serious pathogen in CF is Pseudomonas aeruginosa (Pa).
- Pa lung infections are usually intermittent at first but eventually chronic, leading to mucoid Pa infection usually in the late teens, after which lung function decline is precipitous.
- There is a belief that if Pa infections can be eradicated rapidly, then the mucoid stage can be delayed significantly.
- Our goal is to find the best choice of treatment each time a patient is infected with CF, beginning at birth, to yield the longest mucoid-free survival.

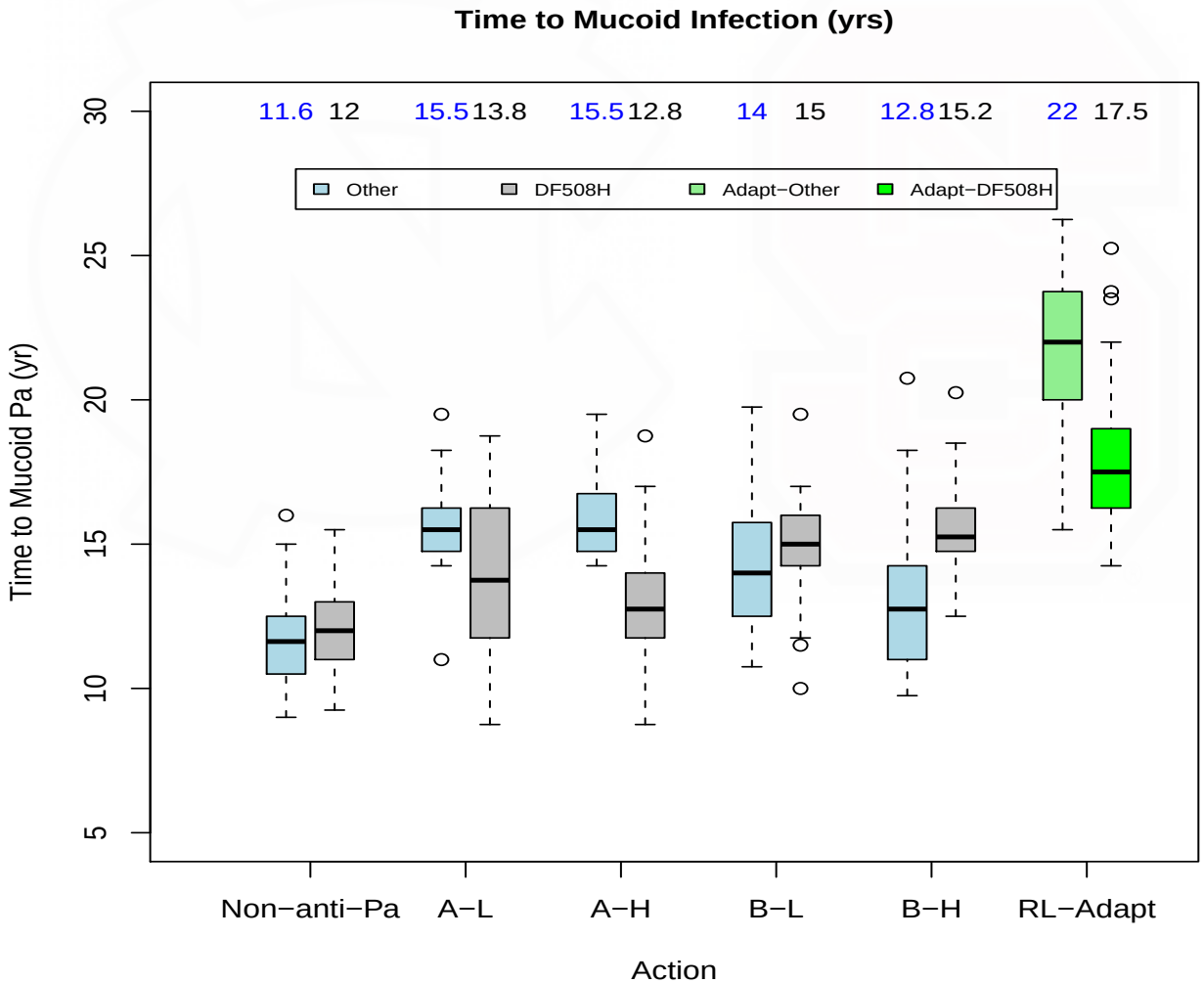
Example 3: Cystic Fibrosis

Realistic simulated patients and trial (Tang, et al., 2012):

- We recruit patients with ages 0–20 years old and follow for about 2 years for Phase II SMART trial.
- For each episode of Pa infection, we randomize to one of 5 treatments: placebo, AL, AH, BL and BH.
- Which treatments are acceptable depends on patient prognostic data, including age.
- After SMART trial completion, we use Q-learning for an "infinite horizon" to estimate optimal, personalized treatment choice as a function of prognostic values.
- A phase III randomize trial is then conducted to verify superiority of the personalized treatment compared to fixed, standard-of-care approaches.

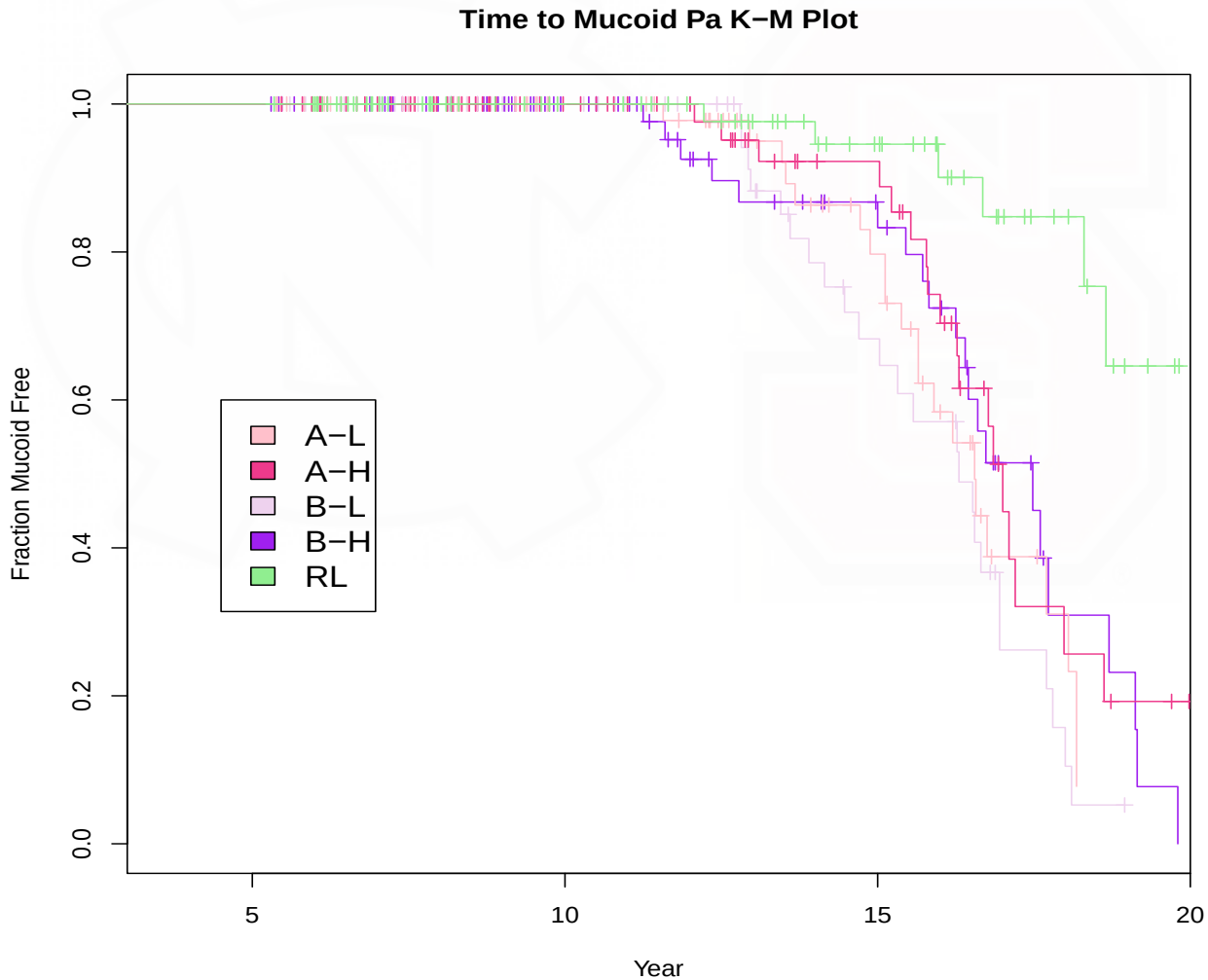
Example 3: Cystic Fibrosis

Comparison of time-to-mucoid infection between optimal personalized treatment and fixed treatments from SMART trial:



Example 3: Cystic Fibrosis

Kaplan-Meier plots from 5 year confirmatory Phase III trial of optimal versus fixed regimens:



Example 3: Cystic Fibrosis

Some scientific and statistical issues:

- Construction of primary clinical outcome (utility) as a composite of several outcomes was highly non-trivial.
- The fact that the disease course is much longer than feasible clinical trial durations raises clinical trial design and treatment regime estimation challenges.
- The way we addressed this:
  - 2-year SMART Phase II trial with variety of ages.
  - 5-year Confirmatory Phase III trial also with variety of ages.
  - Careful selection of utility to include short time outcomes predictive of mucoid PA along as well as mucoid PA.
  - Judicious use of an infinite horizon Q-function which was assumed to be constant across decision times.

Overview of Statistical Issues

- Complex structure of Q-functions
  - Nonlinearity
  - Complicated interactions
  - High dimensional data
- Complex decision making
  - Drug choice
  - Timing of treatment
  - Dose level
- Censoring
- Clinical trial design challenges

Statistical Learning

Statistical learning consists of data driven tools for regression, classification and for other facets of decision making.

Many approaches originated in computer science (artificial intelligence and machine learning) but have more recently become part of statistical science (statistical learning).

Examples include:

- Support vector machines (SVM)
- Support vector regression (SVR)
- Random forests
- Reinforcement learning
- Q-learning and A-learning

Statistical Learning

Advantages

- Good at handling nonlinearity, complicated interactions and high dimensional data
- Computationally efficient with available software
- Can address prediction issues not covered by regression or classification

Disadvantages

- Cannot in general handle censoring
- Almost no inference procedures available
- Requires indirect estimation of decision function $d(x)$ through first estimating $Q(x, a) = E(Y|X = x, A = a)$ and then inverting via $\hat{d}(x) = \text{argmax}_a \hat{Q}(x, a)$.

Statistical Learning

Single decision setting

- Powerful statistical learning tools for estimating $Q(x, a)$ including support vector regression and random forests.
- Ability to handle censoring is almost nonexistent.
- Requires indirect estimation of $d(x)$ via $Q(x, a)$.

Multiple decision setting

- Reinforcement learning enables estimation of decision functions through Q-functions at each decision point using either traditional regression (e.g., linear regression) or statistical learning.
- Censoring and indirect estimation also challenges here.

Incorporating Censoring

Basic issue

The basic issue is that in estimating Q-functions where the outcome Y is a failure time, we are interest in a conditional expectation rather than the more standard hazard function in survival analysis.

Ad hoc approaches

- Censoring is almost never encountered in computer science based artificial intelligence approaches.
- One could throw out the censored observations.
- Another approach for SVR is to not penalize if the prediction is above the censored observation and only penalize if below: this is better than the above but still has significant bias.

Incorporating Censoring

Progress for single decision setting:

- Successfully developed new random forest approach for censored data, "Recursively Imputed Survival Trees" (Zhu and Kosorok, 2012).
- The above approach is very computationally efficient and avoids inverse weighting.
- Extended support vector regression to survival data using inverse probability of censoring weighting (Goldberg and Kosorok, 2012a).
- The above approach is consistent, with good error rates, and performs well, but the inverse weighting requires additional modeling of censoring.

Incorporating Censoring

Progress for multiple decision setting:

- Ad hoc approach based on decreased penalization for censored observations performed reasonably well in two-stage Q-learning for treating non-small cell lung cancer (Zhao, et al., 2011).
- However, theoretically, the above ad hoc approach can potentially have unbounded bias.
- Successfully developed Q-learning for right censored data using inverse probability of censoring weighting (Goldberg and Kosorok, 2012b).
- The approach is known to be asymptotically unbiased with good error rates and is computationally reasonable.

Outcome Weighted Learning

1. Let P denote the distribution of (X, A, Y), where treatments are randomized, and P^d denoted the distribution of (X, A, Y), where treatments are chosen according to d. The value function of d (Qian & Murphy, 2011) is

$$V(d) = E^d(Y) = \int Y dP^d = \int Y \frac{dP^d}{dP} dP = E\left[\frac{I(A = d(X))}{P(A|X)} Y\right].$$

2. Optimal Individualized Treatment Rule:

$$d^* \in \text{argmax}_d V(d).$$

$$E(Y|X, A = 1) > E(Y|X, A = -1) \Rightarrow d^*(X) = 1$$
$$E(Y|X, A = 1) < E(Y|X, A = -1) \Rightarrow d^*(X) = -1$$

Outcome Weighted Learning (OWL)

Optimal Individualized Treatment Rule d^*

Maximize the value $\quad$ Minimize the risk

$$E\left[\frac{I(A = d(X))}{P(A|X)}Y\right] \quad E\left[\frac{I(A \neq d(X))}{P(A|X)}Y\right]$$

- For any rule d, $d(X) = \text{sign}(f(X))$ for some function f.
- Empirical approximation to the risk function:

$$n^{-1}\sum_{i=1}^{n}\frac{Y_i}{P(A_i|X_i)}I(A_i \neq \text{sign}(f(X_i))).$$

- Computation challenges: non-convexity, discontinuity of loss.

Outcome Weighted Support Vector Machine

Objective Function: Regularization Framework

$$\min_f \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{P(A_i|X_i)} \phi(A_i f(X_i)) + \lambda_n \|f\|^2 \right\}. \qquad (2)$$

- $\phi(u) = (1 = u)^+$ is the hinge loss surrogate, $\|f\|$ is some norm for f, and λ_n controls the severity of the penalty on f.
- A linear decision rule: $f(X) = X^T\beta + \beta_0$, with $\|f\|$ as the Euclidean norm of β.
- Estimated individualized treatment rule:

$$\hat{d}_n = \text{sign}(\hat{f}_n(X)),$$

where $\hat{f}_n$ is the solution to (2).

OWL Results

- Fisher consistent and asymptotically consistent.
- Risk bounds and convergence rates similar to those observed in SVM literature (Tsybakov, 2004).
- Excellent simulation results.
- Promising performance overall (Zhao, et al., 2012a).
- Opens door to application of statistical learning techniques to personalized medicine.

OWL: Nefazodone-CBASP Clinical Trial (Keller et al., 2000)

- 681 patients with non-psychotic chronic major depressive disorder (MDD).
- Randomized in a 1:1:1 ratio to either nefazodone, cognitive behavioral-analysis system of psychotherapy (CBASP) or the combination of nefazodone and psychotherapy.
- Primary outcome: score on the 24-item Hamilton Rating Scale for Depression (HRSD); the lower the better.
- 50 baseline variables: demographics, psychological problem diagnostics etc.

OWL: Nefazodone-CBASP Clinical Trial (Keller et al., 2000)

Pairwise Comparison:

- OWL: Gaussian kernel.
  l_1-PLS and OLS: $(1, X, A, XA)$.
- Value calculated with a 5-fold cross validation type analysis.

Table: Mean HRSD (Lower is Better) from Cross Validation Procedure with Different Methods

	OLS	l_1-PLS	OWL
Nefazodone vs CBASP	15.87	15.95	15.74
Combination vs Nefazodone	11.75	11.28	10.71
Combination vs CBASP	12.22	10.97	10.86

OWL: Comments

The Outcome Weighted Learning procedure

- Discovers an optimal individualized therapy to improve expected outcome.
- Nonparametric approach sidesteps the inversion step and invokes statistical learning techniques directly.

Some open questions:

- How to handle censoring?
- How to generate sample size formulas to enable practical Phase II design?
- How to handle deciding among more than two treatments?

OWL for Multiple-Stage Decision Making

Problems with Q learning

- Mismatch exists between estimating the optimal Q function and the goal of maximizing the value function (Murphy, 2005).
- Non-smooth maximization operation.
- High dimensional covariate space.

Backwards Outcome Weighted Learning (BOWL)

- Generalization of OWL to multi-decision setup (Zhao, et al., 2012b).
- Find the optimal decision rule by directly maximizing the value function for each stage backwards repeatedly.
- Consistency and risk bound of BOWL estimator.

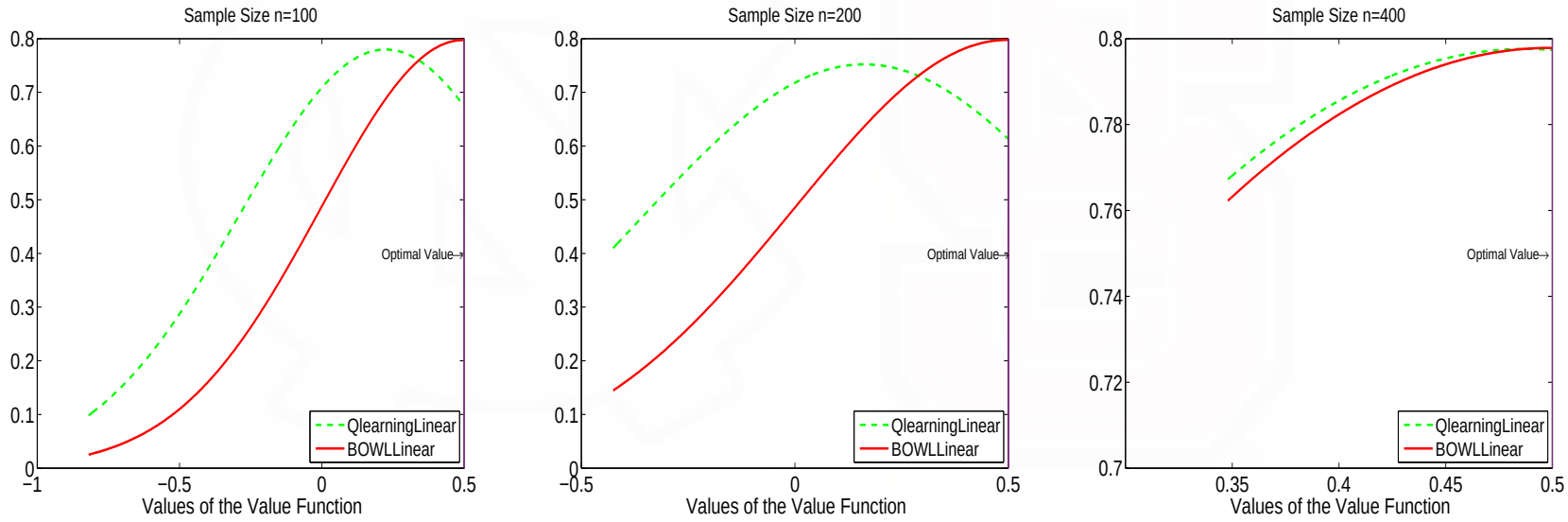
BOWL: Simulation Study

Generative Model (Chakraborty et al., 2010)

- $X_1 \sim U[-1,1]^{50}$, $X_2 = X_1$.
- $A_1, A_2 \in \{-1, 1\}, P(A_1 = 1) = P(A_2 = 1) = 0.5$.
- $Y_1 = 0$, $Y_2 | H_2, A_2 \sim N(-0.5A_1 + 0.5A_2 + 0.5A_1A_2, 1)$.

- Training data sample size $n = 100, 200, 400$.
- Testing data sample size 10000.
- 500 replications.
- Methods: BOWL with Linear kernel; Q learning with linear regression.

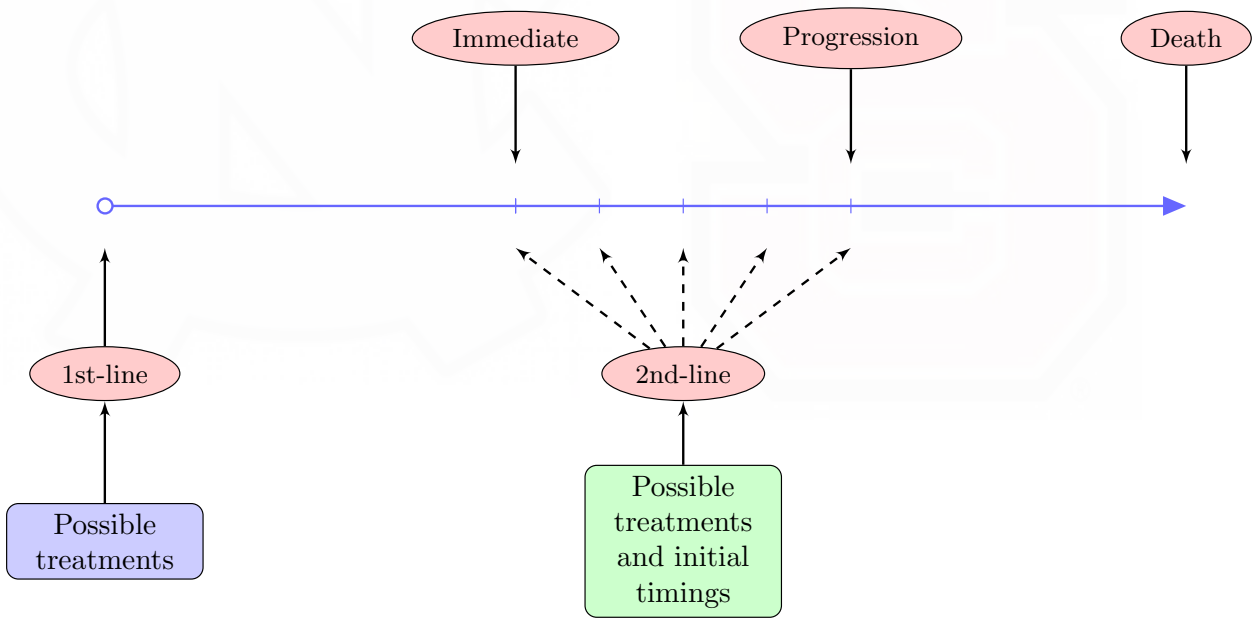
BOWL: Simulation Results



Note: Q learning encounters difficulties with small sample sizes.

Open Questions for OWL and BOWL

- Survival outcomes
- Multicategory/Continuous treatments.
  - Multiple therapies.
  - Continuous range of dose levels.
- Optimize timing to switch treatments in multi-stage trials.



Other Open Questions

- Development of meaningful inference tools: this is hard even for linear regression in Q-learning.
- Develop sample size algorithms or formulas.
- When should parametric or semiparametric approaches be used instead of machine learning approaches?
- How to design trials for long-term chronic diseases.
- How to elicit and formulate outcomes (utility).
- How to handle continuing reassessment so that previously developed regimes could be enlarged to include new and emerging treatments.

Preparing Protocols

- Each setting seems to be unique.
- Often best to frame the trial first as a traditional trial with randomized treatments and then add personalized medicine and dynamic treatment regime aspect as later aims.
- There are ways to frame dynamic treatment regime estimation, in some cases, as weighted linear regression.
- Sample sizes roughly correspond to large traditional Phase II (or small Phase III) designs for SMART trials.
- We are working on sample size software for OWL studies.
- We have completed or are working on about 5 such trials.

References

- Chakraborty, B., Murphy, S. A., and Strecher, V. (2009). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* 19:317–343.
- Goldberg, Y., and Kosorok, M. R. (2012a). Support Vector Regression for Right Censored Data. Submitted.
- Goldberg, Y., and Kosorok, M. R. (2012b). Q-learning with censored data. *Annals of Statistics* 40:529–560.
- Qian, M., and Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *Annals of Statistics* 39:1180–1210.
- Tang, Y., and Kosorok, M. R. (2012). Developing adaptive personalized therapy for cystic fibrosis using reinforcement learning. UNC-Chapel Hill Biostatistics Technical Report 30.

References, cont'd

- Tsybakov, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Annals of Statistics* 32:135–166.
- Zhao, Y., Zeng, D., Socinski, M. A., and Kosorok, M. R. (2011). Reinforcement learning strategies for clinical trials in non-small cell lung cancer. *Biometrics* 67:1422–1433.
- Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012a). Estimating individualized treatment rules using outcome weighted learning. *JASA* 107:1106-1118.
- Zhao, Y., Zeng, D., Laber, E. B., and Kosorok, M. R. (2012b). New statistical learning methods for estimating optimal dynamic treatment regimes. Submitted.
- Zhu, R., and Kosorok, M. R. (2012). Recursively imputed survival trees. *JASA* 107:331–340.