Section V: Extension

- Survival outcomes with censoring
- Multicategory Treatment
- Dynamic Treatment Regimes
- Concluding Remark on Observational Studies

Survival outcomes with censoring

Survival Outcomes with Censoring

- Interested in time-to-event outcome.

- Observe independently and identically distributed training data $(X_i, A_i, D_i, \Omega_i), i = 1, \ldots, n$.

  X: baseline variables, $X \in \mathbb{R}^p$,
  A: binary treatment options, $A \in \{0, 1\}$,
  D: observed event time.
  Ω: censoring indicator $\Omega_i = I(T_i \leq C_i)$.

- $D = \min(T, C)$: T survival time, C censoring time.

- Throughout assume:
  - Conditionally independent censoring: $T \perp C \mid A, X$
  - Positivity: within all strata of (A, X), there is a positive probability of observing individuals until the end of the study (or, until the end of the time window of interest)

Survival Outcomes with Censoring

- We focus on two possible objectives:
  - Maximize the probability of surviving beyond a landmark time;
  - Maximize restricted mean survival time.
- The different objectives can lead to different optimal rules
- Other objective functions may also be interesting – e.g., those that incorporate quality-of-life measures

Probability of surviving beyond a landmark time

Let T be the event time. Let $D = I(T < t_0)$ be an indicator that the event occurs before a landmark time t_0.

- Estimate $E(D|A, X)$ using a regression method suitable for time-to-event outcomes, e.g., Cox regression with treatment-by-covariate interactions (Cox 1972) or doubly robust approaches (e.g., Rubin and van der Laan, 2007).

- Consider performing analyses for different choices of t_0; often X more weakly predicts treatment effect for larger t_0.

Cox, *JRSSB*, 1972; Rubin & van der Laan, *Int J Biostat*, 2007

Restricted mean survival time

- Regression modeling approach: inverse probability of censoring weighted (IPW) Q-learning:
  - $E(D|A,X)$ is modeled using treatment-by-covariate interactions, accounting for the probability of being censored.
- Outcome weighted learning approach:
  - Replace D_i by $\Omega_i D_i / \hat{S}_C(D_i|A_i,X_i)$ in the outcome weighted learning for uncensored data, where $\widehat{S}_C(D|A,X)$ is the estimated conditional survival function of C given (A,X).
- Can also extend the approach that directly estimates the contrast function $\Delta(X) = E[D|A=0,X] - E[D|A=1,X]$
  - Simply need to replace the outcome D_i by the inverse probability of censoring weighted outcome $\Omega_i D_i / \hat{S}_C(D_i|A_i,X_i)$

Goldberg & Kosorok (*Annals of Stat.*, 2012); Zhao et al (*Biometrika*, 2015)

Evaluation in the censoring data setup

- Estimate performance measures empirically using inverse-probability-of-censoring weights or doubly robust approaches (e.g., Bang and Robins, *Biometrics*, 2005)

- Model-based estimates require no modification.

Multicategory Treatment

Multicategory Treatment

- More than two treatments of interest: $A \in \{1, \ldots, K\}$
  - e.g., $K = 3$ in depression data
- $d^*(x) = \text{argmin}_{k=1,\ldots,K}\, \mu(k, x)$.
- Can estimate $E(D|A, X)$ with $\hat{\mu}(A, X)$
- The estimator for the optimal treatment regime

$$\hat{d}_n(x) = \underset{k=1,\ldots,K}{\text{argmin}}\, \hat{\mu}(k, x).$$

- Contrast function estimation valid with almost no modification if can choose a single reference treatment assignment k^{ref} and subsequently define

$$\Delta(k, X) = E[D|A = k^{ref}, X] - E[D|A = k, X]$$

.

Dynamic Treatment Regimes

Dynamic Treatment Regimes (DTRs)

- Motivation: treatment of chronic illnesses
  - Some examples: HIV/AIDS, cancer, depression, schizophrenia, drug and alcohol addiction, ADHD, etc.
  - Multistage decision making problem
  - Longer-term treatment requires consideration and tradeoff between immediate and longer term benefit.

Murphy, *JRSS-B* (2003)

Dynamic Treatment Regimes

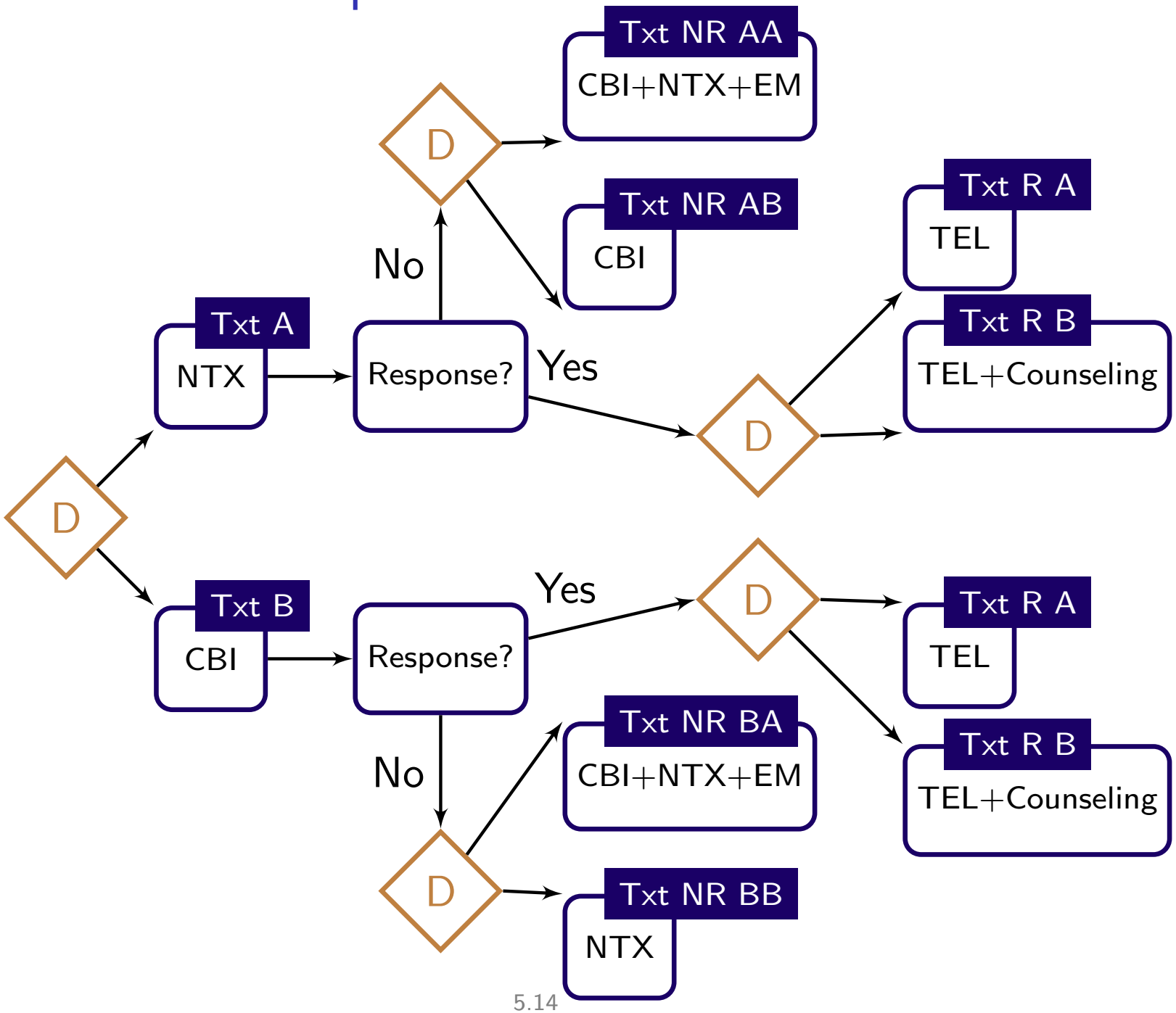
- Operationalize multistage decision making via as sequence of decision rules
  - One decision rule for each time (decision) point
  - A decision rule is a function inputs patient history and outputs a recommended treatment
- Aim to optimize some cumulative clinical outcome
  - Survival time
  - Depression test scores
  - Indicator of no myocardial infarction within 30 days ...

Dramatized Example

- Addiction management example inspired by the ExTENd and COMBINE trials (Murphy et al, 2007)

- Devising two-time point treatment strategy for alcohol dependent patients.
  - Initial treatment choices Naltrexone (NTX) and Combined Behavioral Intervention (CBI).
  - At six-months responders classified as responders or non-responders.
  - For responders to initial treatment, followup treatment choices are telephone monitoring (TEL) and telephone monitoring + counseling (TEL+Counseling).
  - For non-responders to initial treatment, followup treatment choices are switch initial treatments (NTX $\leftrightarrow$ CBI), or step-up initial treatment CBI + NTX + Enhanced monitoring (CBI + NTX +EM).

Dramatized Example



Dramatized Example

- H_j denote history at stage j.
- At presentation: Baseline variables x_1; accrued information $h_1 = x_1$
  - Decision point 1: Two treatment options {NTX, CBI}; rule 1: $d_1(h_1) \Rightarrow d_1 : h_1 \rightarrow \{\text{NTX, CBI}\}$
  - Between decisions 1 and 2: Collect additional information x_2, including responder status
  - Accrued information $h_2 = \{x_1, \text{treatment at decision 1}, x_2\}$
  - Decision point 2: Four options

Optimal Dynamic Treatment Regimes

- Examples of treatment regimes: Prescribe NTX initially; then assign TEL to responders; and assign step-up to non-responders.
- Optimal DTR d^* leads to the lowest expected outcome among all possible regimes

Challenges in Estimating Optimal DTRs: Delayed Effects

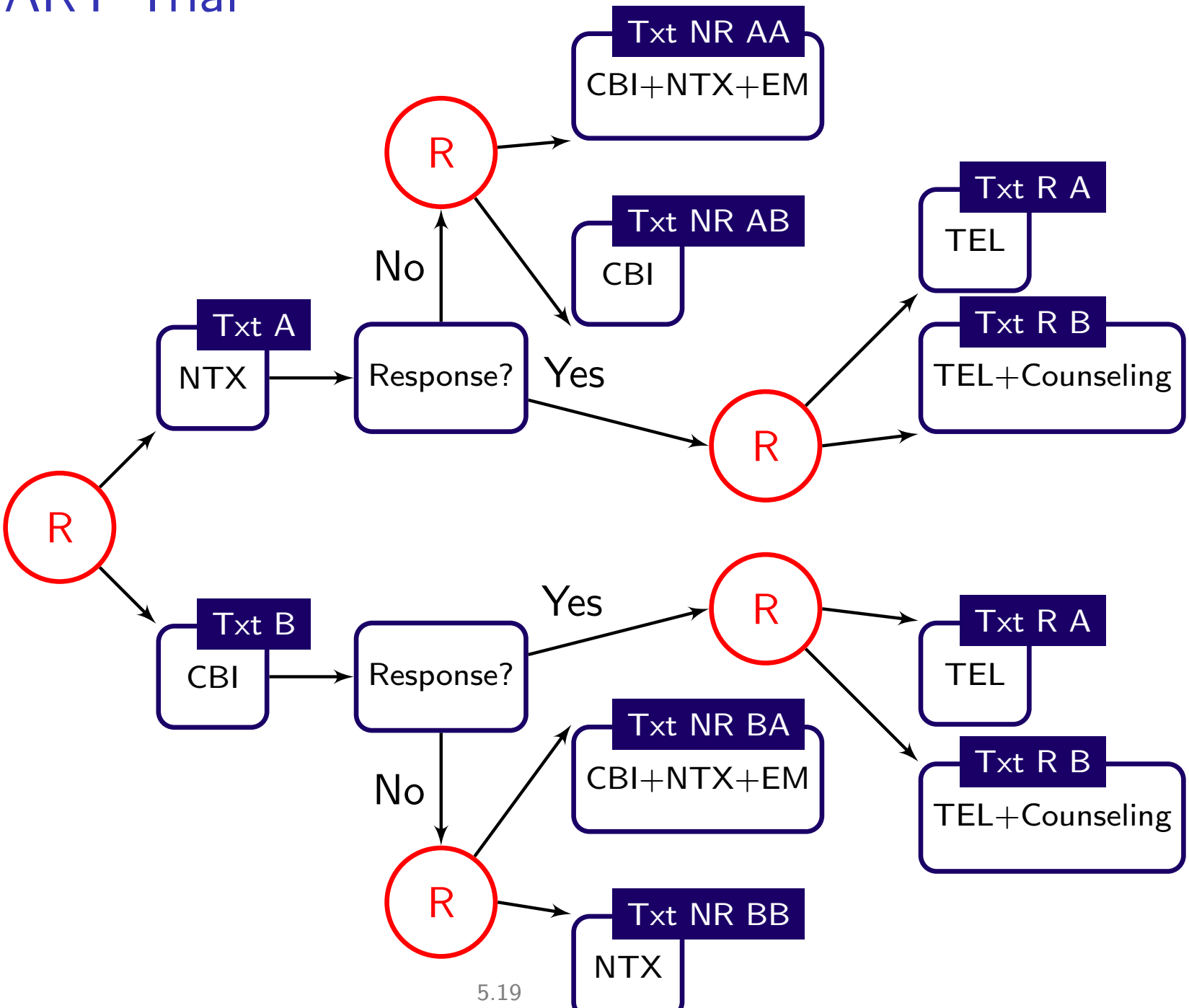
- The therapy with the higher proportion of responders might have other effects that render subsequent treatments less effective in regard to the final response.
- The therapy with lower proportion of responders may not appear best initially but may have enhanced long term effectiveness when followed by a particular maintenance treatment.
- Must consider the entire sequence of decisions
- Must accommodate intermediate information including prior treatments into current treatment choice.

Sequential Multiple Assignment Randomized Trial (SMART)

- Due the the aforementioned challenges, it would be ideal to adopt a particular design to best estimate the optimal DTRs
- SMART: designed for estimation of optimal DTRs
- Randomize subjects to the treatment options at each decision point
- Take advantage of sequential randomization to eliminate confounding
- Collect both initial and intermediate information on possible tailoring variables

Murphy (*Stat in Med*, 2005)

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

Data

- (X_1, A_1, X_2, A_2, D) for each individual
  X_k: Observations available at stage k
  A_k: Treatment at stage k
  D: Primary outcome
  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 the lowest $E^d(D)$, the expected outcome if all patients are assigned treatment according to d

Dynamic Programming

- Optimal regime d^* can be derived using dynamic programming (Bellman, 1957)

  1. For a given history at the final time point, define the expected mean outcome: $Q_2(h_2, a_2) = E\left(D \middle| H_2 = h_2, A_2 = a_2\right)$

  2. Let $d_2^*(h_2)$ denote the treatment decision that minimizes the mean outcome given the history h_2, i.e. $d_2^*(h_2) = \arg\min_{a_2 \in \{0,1\}} Q_2(h_2, a_2)$

  3. At the first time point, the goal is to assign the best treatment given that d_2^* will be followed in the second time point. So, the expected outcome under treatment assignment a_1 given history h_1 is:

$$Q_1(h_1, a_1) = E\left[Q_2(H_2, d_2^*(H_2)) | H_1 = h_1, A_1 = a_1\right]$$

  4. Define $d_1^*(h_1)$ as the treatment decision that minimizes the mean outcome at time 1 given history h_1, i.e. $d_1^*(h_1) = \arg\min_{a_1 \in \{0,1\}} Q_1(h_1, a_1)$

Dynamic Programming as a Series of Single Time Point Decisions

The dynamic programming scheme on the last slide allows us to break the problem into a series of single time point treatment decisions:

- Initialize $D_{K+1} = D$. (K is the number of time points at which tx decisions are made)
- For $t = \{K, K-1, \ldots, 1\}$
  1. Let $Q_t(h_t, a_t) = E[D_{t+1} | A_t = a_t, H_t = h_t]$.
  2. Let $d_t(h_t) = \text{argmin}_{a_t \in \{0,1\}} Q_t(h_t, a_t)$.
  3. For each individual with history H_t, let $D_t = Q_t(H_t, d_t(H_t))$.

The optimal rule definition in Step 2 has the same form as the rules that we've studied throughout this course.

Consequently, everything we have learned in this course can now be used in a recursive form to learn optimal treatment rules (the next slide specializes this observation to Q-learning).

Constructing a DTR from Data: Q-learning

- When system dynamics are known dynamic programming yields the optimal DTR, but we only have data
- Q-learning: data-driven analog of dynamic programming: replaces conditional expectations with regression models
- Recursively estimates the Q-function, starting at the final time point and progressing backwards in time.
- The estimated optimal sequence of decision rules

$$\hat{d}_j(h_j) = \underset{a_j \in \{0,1\}}{\text{argmin}}\, \hat{Q}_j(h_j, a_j).$$

- An extension of regression to sequential treatments.

Summary

- Data from SMART designs can be used to construct optimal DTRs
- Q learning is a common method, though it has some drawbacks, e.g., require correct specified models
- Many other methods have been developed.

Concluding Remark on Observational Studies

We close by recalling that nearly everything that we discussed during this course can also be applied in observational studies, **under the following conditions**:

- Positivity: $P(A = a|X = x)$ strictly positive for all x
- Consistency: $D(a) = D$ whenever treatment a is actually received
- No unmeasured confounders:

$$D(0) \perp A|X \qquad \text{and} \qquad D(1) \perp A|X$$

  - X contains all information used to assign treatments