Implementing Personalized Medicine: Estimating Optimal Dynamic Treatment Regimes

Baqun Zhang, Phillip Schulte, Anastasios Tsiatis, Eric Laber, and Marie Davidian

Department of Statistics
North Carolina State University

http://www4.stat.ncsu.edu/~davidian

Dynamic Treatment Regime

Clinical decision-making: Clinicians make *treatment decisions* over the course of a patient's disease or disorder

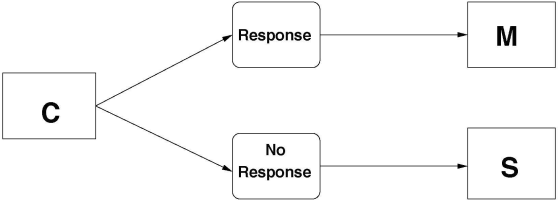
- Key *decision points*, multiple *treatment options* at each

Formalizing clinical decision-making: At any *decision*

- Need a *rule* that *inputs* the *accrued information* on the patient to that point and *outputs* the next treatment from among the *available options*

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

Butch's Cancer Example



Two decision points:

- *Decision 1*: Induction chemotherapy (options c_1, c_2)
- *Decision 2*:
  - Maintenance treatment for patients who *respond* (options m_1, m_2)
  - Salvage chemotherapy for those who *don't* (options s_1, s_2)

Summary

K key decision points:

- *Baseline information* x_1, *intermediate information* x_k between decisions $k-1$ and k, $k = 2, \ldots, K$
- Set of *treatment options* at each decision k: $a_k \in \mathcal{A}_k$
- *Accrued information* $h_1 = x_1 \in \mathcal{H}_1$,

$$h_k = \{x_1, a_1, x_2, a_2, \ldots, x_{k-1}, a_{k-1}, x_k\} \in \mathcal{H}_k, \quad k = 2, \ldots, K$$

- *Decision rules* $d_1(h_1), d_2(h_2), \ldots, d_K(h_K)$, $d_k : \mathcal{H}_k \rightarrow \mathcal{A}_k$
- *Treatment regime* $d = (d_1, d_2, \ldots, d_K)$

Optimal Treatment Regime

“Best” treatment regime:

- $\mathcal{D}$ = class of *all possible* dynamic treatment regimes
- *Clinical outcome* such that *larger outcomes* are *better* (e.g., survival time)
- *Potential outcome* $Y^*(d)$ = outcome a patient *would achieve* if s/he received treatment *according to* $d \in \mathcal{D}$
- $E\{Y^*(d)\}$ = expected outcome for the *population* if *all patients* received treatment *according to* $d \in \mathcal{D}$
- *Optimal regime* d^{opt} *maximizes*

$$E\{Y^*(d)\} \quad \text{among all } \; d \in \mathcal{D}$$

Goal: Under *suitable assumptions*, *estimate*

$$d^{opt} = (d_1^{opt}, \ldots, d_K^{opt}) \quad \text{based on } data$$

Single Decision

Single decision, $K = 1$, 2 treatment options: *Data*

$$(X_{1i}, A_{1i}, Y_i), \;\; i = 1, \ldots, n, \quad \mathcal{A}_1 = \{0, 1\}$$

Key assumption: *No unmeasured confounders*

- $Y^*(0)$, $Y^*(1)$ are *potential outcomes* under treatments 0, 1
- With $d(x_1)$ taking values 0, 1

$$Y^*(d) = Y^*(1)d(X_1) + Y^*(0)\{1 - d(X_1)\}$$

- *No unmeasured confounders*

$$Y^*(0), Y^*(1) \perp\!\!\!\perp A_1 \mid X_1$$

- Automatically satisfied in a *clinical trial*
- *Unverifiable* but necessary in an *observational study*

Critical Result

Under no unmeasured confounders:

$$
\begin{aligned}
E\{Y^*(1)\} &= E[E\{Y^*(1)|X_1\}] = E[E\{Y^*(1)|X_1, A=1\}] \\
&= E\{\, E(Y|X_1, A=1)\,\}
\end{aligned}
$$

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

Can be used to show that:

$$
\begin{aligned}
E\{Y^*(d)\} &= E[E\{Y^*(d)|X\}] \\
&= E\Big[E\{Y^*(1)|X_1\}d(X_1) + E\{Y^*(0)|X_1\}\{1-d(X_1)\}\Big] \\
&= E\Big[\, E(Y|X_1, A=1)d(X_1) + E(Y|X_1, A=0)\{1-d(X_1)\}\,\Big]
\end{aligned}
$$

Estimating d^{opt}

$$E\{Y^*(d)\} = E\Big[E(Y|X_1, A=1)d(X_1) + E(Y|X_1, A=0)\{1 - d(X_1)\} \Big]$$

Thus: With $Q(x_1, a_1) = E(Y|X_1 = x_1, A_1 = a_1)$

$$d^{opt}(x_1) = I\{Q(x_1, 1) > Q(x_1, 0)\}$$

Regression estimator for d^{opt}:

- Posit and fit a *regression model* $Q(x_1, a_1; \beta)$
- *Estimate* d^{opt} by

$$\widehat{d}^{opt}(x_1) = I\{\, Q(x_1, 1; \widehat{\beta}) > Q(x_1, 0; \widehat{\beta})\, \}$$

Estimating d^{opt}

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

Alternative method: *Value search*

- Specify a *class* of regimes $\mathcal{D}_\eta \in \mathcal{D}$, e.g.,

$$d(x_1; \eta) = d_\eta(x_1) = I(x_{1,1} < \eta_0, x_{1,2} < \eta_1)$$

- $\mathcal{D}_\eta$ chosen based on *feasibility*, *interpretability*, or *cost*
- $d_\eta^{opt}(x_1) = d(x_1; \eta^{opt})$, η^{opt} *maximizes* $E\{Y^*(d_\eta)\}$ in η
- *Suggests*: *Maximize* an estimator for $E\{Y^*(d_\eta)\}$ in η

Value search estimator for d^{opt}:

- *Missing data analogy*: For fixed η, define

$$C_\eta = A_1 d(X_1;\eta) + (1 - A_1)\{1 - d(X_1;\eta)\}$$

- $C_\eta = 1$ if the treatment received is *consistent with* having followed d_η and = 0 otherwise
- Only subjects with $C_\eta = 1$ have observed outcomes *consistent with following* d_η; for others, these are *missing*
- *Propensity score* $\pi(X_1) = \text{pr}(A_1 = 1|X_1)$ (*known* in a trial; otherwise must model and fit)
- *Propensity* of receiving treatment *consistent with* d_η

$$\begin{aligned}
\pi_c(X_1;\eta) &= \text{pr}(C_\eta = 1|X_1) \\
&= \pi(X_1)d(X_1;\eta) + \{1 - \pi(X_1)\}\{1 - d(X_1;\eta)\}
\end{aligned}$$

Estimating d^{opt}

Inverse probability weighted estimator for $E\{Y^*(d_\eta)\}$:

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

- *Consistent* for $E\{Y^*(d_\eta)\}$ if $\pi(X_1)$ and hence $\pi_c(X_1;\eta)$ is *correctly specified*
- Find $\widehat{\eta}^{opt}$ maximizing $IPWE(\eta)$ in η

$$\widehat{d}_\eta^{opt}(x_1) = d(x_1;\widehat{\eta}^{opt})$$

- *Better*: *Doubly robust augmented* inverse probability weighted estimator

Multiple Decisions

Same ideas, only more complicated...

Potential outcomes under a regime $d \in \mathcal{D}$:

- *Initial information* X_1, *potential outcomes*

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

Optimal regime: d^{opt} satisfies

$$E\{Y^*(d)\} \leq E\{Y^*(d^{opt})\} \quad \text{for all } d \in \mathcal{D}$$

Estimating d^{opt}

Complication:

- Can't we just estimate the rules at each decision *separately* using previous methods and "*piece together*" the estimated regime from separate studies?
- *Unfortunately not*
- *Delayed effects*: E.g., c_1 may not appear best initially but may have enhanced effectiveness when followed by m_1
- $\Longrightarrow$ Must use data from a *single study* (same patients) reflecting the *entire sequence of decisions* and use methods that *acknowledge* this

Observed Data

Data required: $(X_{1i}, A_{1i}, X_{2i}, A_{2i}, \ldots, X_{(K-1)i}, A_{(K-1)i}, X_{Ki}, Y_i)$, $i = 1, \ldots, n$, iid

- X_1 = *Baseline covariate information*
- X_k, $k = 2, \ldots, K$, = *intermediate information* observed between decisions $k-1$ and k
- A_k, $k = 1, \ldots, K$, = *treatment received* at decision k
- $H_k = (X_1, A_1, X_2, \ldots, A_{k-1}, X_k)$, $k = 2, \ldots, K$, = *accrued information* at decision k
- Y = *observed outcome*; can be *ascertained after* decision K or can be a *function* of $X_2, \ldots, X_K$

Data Sources

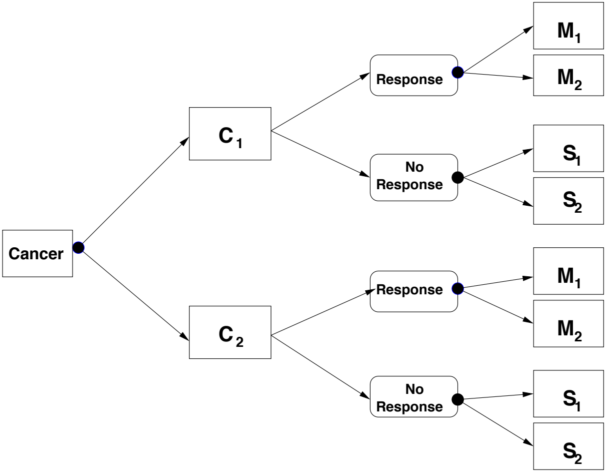
Critical assumption: *Sequential randomization*

- Generalization of *no unmeasured confounders*
- Treatment received at *each decision* $k = 1, \dots, K$ is *statistically independent* of potential outcomes that would be achieved under all treatment options *at all decisions* given the *accrued information* H_k

Studies:

- Longitudinal observational
- *SMART*: Sequential, Multiple Assignment, Randomized Trial

Cancer example: Randomization at •s



Sequential Regression Estimator

Sequential regression: Using *backward induction* – illustrate for $K = 2$ and treatment options $\{0, 1\}$ at each decision

- With $H_2 = (X_1, A_1, X_2)$, *posit and fit a regression model*

$$Q_2(h_2, a_2) = E(Y|H_2 = h_2, A_2 = a_2) \quad \text{and substitute in}$$

$$d_2^{opt}(h_2) = I\{Q_2(h_2, 1) > Q_2(h_2, 0)\}$$

- Move *backward*: *posit and fit a regression model* for expected outcome *assuming* $d_2^{opt}(h_2)$ is used to determine treatment at decision 2 *in the future*, i.e.

$$Q_1(x_1, a_1) = E[\max\{Q_2(H_2, 0), Q_2(H_2, 1)\}|X_1 = x_1, A_1 = a_1]$$

and substitute in

$$d_1^{opt}(x_1) = I\{Q_1(x_1, 1) > Q_1(x_1, 0)]\}$$

- *Q-learning*, form of *dynamic programming*, *reinforcement learning* in computer science

Other Estimators

Alternatives to Q-learning:

- Concern over *model misspecification* with Q-learning
- *A-learning*: model only *contrasts* among treatments
- Extension of *value search* estimator using *dropout analogy*

Challenges

- Estimation of optimal treatment regimes is a *wide open* area of research
- How to deal with *high-dimensional* information? Regression *model selection*?
- "*Black box*" vs. *class* of regimes?
- *Competing outcomes*, e.g., efficacy vs. toxicity
- *Design* of SMARTs?

References

Schulte, P. J., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science*, in press.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. *Biometrics* **68**, 1010–1018.

Zhang, B., Tsiatis, A.A., Davidian, M., Zhang, M., and Laber, E.B., (2012). Estimating optimal treatment regimes from a classification perspective. *Stat* **1**, 103–114.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2013). Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika* **100**, 681–694.