

On Testing Conditional Qualitative Treatment Effects

Chengchun Shi

Department of Statistics
North Carolina State University

Joint work with Wenbin Lu and Rui Song

July 30, 2017

A few words on causal inference

Data

- A : Treatment (0 or 1)
- X : Covariates
- Y : Observed outcome (usually the larger the better)

A few words on causal inference

Data

- A : Treatment (0 or 1)
- X : Covariates
- Y : Observed outcome (usually the larger the better)
- $Y^*(a)$: Potential outcome $a = 0, 1$

A few words on causal inference

Data

- A : Treatment (0 or 1)
- X : Covariates
- Y : Observed outcome (usually the larger the better)
- $Y^*(a)$: Potential outcome $a = 0, 1$

Objective

Identify the optimal regime d_{opt} to reach the best clinical outcome

A few words on causal inference

Data

- A : Treatment (0 or 1)
- X : Covariates
- Y : Observed outcome (usually the larger the better)
- $Y^*(a)$: Potential outcome $a = 0, 1$

Objective

Identify the optimal regime d_{opt} to reach the best clinical outcome

- Maximize $EY^*(d) = E[d(X)Y^*(1) + \{1 - d(X)\}Y^*(0)]$

$$d : X \rightarrow \{0, 1\}.$$

Q, Contrast and Value function

- $Q(x, a) = E[Y|X = x, A = a]$,

Q, Contrast and Value function

- $Q(x, a) = E[Y|X = x, A = a]$,
- $\tau_0(x) = Q(x, 1) - Q(x, 0)$,

Q, Contrast and Value function

- $Q(x, a) = E[Y|X = x, A = a]$,
- $\tau_0(x) = Q(x, 1) - Q(x, 0)$,
- $V(d) = EY^*(d) = E[d(X)Y^*(1) + \{1 - d(X)\}Y^*(0)]$.

Q, Contrast and Value function

- $Q(x, a) = E[Y|X = x, A = a]$,
- $\tau_0(x) = Q(x, 1) - Q(x, 0)$,
- $V(d) = EY^*(d) = E[d(X)Y^*(1) + \{1 - d(X)\}Y^*(0)]$.

Optimal treatment regime

- SUTVA, no unmeasured confounders

Q, Contrast and Value function

- $Q(x, a) = E[Y|X = x, A = a]$,
- $\tau_0(x) = Q(x, 1) - Q(x, 0)$,
- $V(d) = EY^*(d) = E[d(X)Y^*(1) + \{1 - d(X)\}Y^*(0)]$.

Optimal treatment regime

- SUTVA, no unmeasured confounders
- optimal treatment regime

$$d_{opt}(x) = I(\tau_0(x) \geq 0).$$

- There are two types of clinically “important” variables.

- There are two types of clinically “important” variables.
 - **Predictive** variables are those involved in $\tau_0(x)$.

- There are two types of clinically “important” variables.
 - **Predictive** variables are those involved in $\tau_0(x)$.
 - **Prescriptive** variables are those involved in $d_{opt}(x)$.

- There are two types of clinically “important” variables.
 - **Predictive** variables are those involved in $\tau_0(x)$.
 - **Prescriptive** variables are those involved in $d_{opt}(x)$.
- **Predictive** variables have interactions with the treatment.

- There are two types of clinically “important” variables.
 - **Predictive** variables are those involved in $\tau_0(x)$.
 - **Prescriptive** variables are those involved in $d_{opt}(x)$.
- **Predictive** variables have interactions with the treatment.
- **Prescriptive** variables have qualitative interactions with the treatment.

- There are two types of clinically “important” variables.
 - **Predictive** variables are those involved in $\tau_0(x)$.
 - **Prescriptive** variables are those involved in $d_{opt}(x)$.
- **Predictive** variables have interactions with the treatment.
- **Prescriptive** variables have qualitative interactions with the treatment.
- **Prescriptive** variables \subseteq **predictive** variables.

- There are two types of clinically “important” variables.
 - Predictive variables are those involved in $\tau_0(x)$.
 - Prescriptive variables are those involved in $d_{opt}(x)$.
- Predictive variables have interactions with the treatment.
- Prescriptive variables have qualitative interactions with the treatment.
- Prescriptive variables \subseteq predictive variables.
- Predictive variables $\not\subseteq$ prescriptive variables.

- A tiny example:

$$\tau_0(x) = \exp(-x_1)x_2,$$

for $\{x_1, x_2, \dots, x_p\} \in [-1, 1]^p$.

- A tiny example:

$$\tau_0(x) = \exp(-x_1)x_2,$$

for $\{x_1, x_2, \dots, x_p\} \in [-1, 1]^p$.

- $d_{opt}(x) = I\{\exp(-x_1)x_2 \geq 0\} = I(x_2 \geq 0)$.

- A tiny example:

$$\tau_0(x) = \exp(-x_1)x_2,$$

for $\{x_1, x_2, \dots, x_p\} \in [-1, 1]^p$.

- $d_{opt}(x) = I\{\exp(-x_1)x_2 \geq 0\} = I(x_2 \geq 0)$.
- x_1 and x_2 are the **predictive** variables.

- A tiny example:

$$\tau_0(x) = \exp(-x_1)x_2,$$

for $\{x_1, x_2, \dots, x_p\} \in [-1, 1]^p$.

- $d_{opt}(x) = I\{\exp(-x_1)x_2 \geq 0\} = I(x_2 \geq 0)$.
- x_1 and x_2 are the **predictive** variables.
- x_2 is the **prescriptive** variable.

- A tiny example:

$$\tau_0(x) = \exp(-x_1)x_2,$$

for $\{x_1, x_2, \dots, x_p\} \in [-1, 1]^p$.

- $d_{opt}(x) = I\{\exp(-x_1)x_2 \geq 0\} = I(x_2 \geq 0)$.
- x_1 and x_2 are the **predictive** variables.
- x_2 is the **prescriptive** variable.
- x_1 and x_2 have interactions with the treatment.

- A tiny example:

$$\tau_0(x) = \exp(-x_1)x_2,$$

for $\{x_1, x_2, \dots, x_p\} \in [-1, 1]^p$.

- $d_{opt}(x) = I\{\exp(-x_1)x_2 \geq 0\} = I(x_2 \geq 0)$.
- x_1 and x_2 are the **predictive** variables.
- x_2 is the **prescriptive** variable.
- x_1 and x_2 have interactions with the treatment.
- x_2 has **qualitative** interaction with the treatment.

- Gunter et al. (2011) proposed an *S*-score method for quantifying the magnitude of the qualitative treatment effects.

- Gunter et al. (2011) proposed an S -score method for quantifying the magnitude of the qualitative treatment effects.
- Chang et al. (2015) and Hsu (2017) proposed nonparametric tests for testing the qualitative treatment effects.

- Gunter et al. (2011) proposed an S -score method for quantifying the magnitude of the qualitative treatment effects.
- Chang et al. (2015) and Hsu (2017) proposed nonparametric tests for testing the qualitative treatment effects.
- Here, we focus on the **conditional** qualitative treatment effects (CQTE).

- Gunter et al. (2011) proposed an S -score method for quantifying the magnitude of the qualitative treatment effects.
- Chang et al. (2015) and Hsu (2017) proposed nonparametric tests for testing the qualitative treatment effects.
- Here, we focus on the **conditional** qualitative treatment effects (CQTE).
 - Formalize the notion of CQTE and present equivalent representations of no CQTE.

- Gunter et al. (2011) proposed an S -score method for quantifying the magnitude of the qualitative treatment effects.
- Chang et al. (2015) and Hsu (2017) proposed nonparametric tests for testing the qualitative treatment effects.
- Here, we focus on the **conditional** qualitative treatment effects (CQTE).
 - Formalize the notion of CQTE and present equivalent representations of no CQTE.
 - Propose a testing procedure for testing the existence of CQTE.

- Gunter et al. (2011) proposed an S -score method for quantifying the magnitude of the qualitative treatment effects.
- Chang et al. (2015) and Hsu (2017) proposed nonparametric tests for testing the qualitative treatment effects.
- Here, we focus on the **conditional** qualitative treatment effects (CQTE).
 - Formalize the notion of CQTE and present equivalent representations of no CQTE.
 - Propose a testing procedure for testing the existence of CQTE.
 - Develop a variable selection procedure for selecting **prescriptive** variables in a sequential order.

Conditional qualitative treatment effects (CQTE)

- B and C are two disjoint subsets of $[1, \dots, p]$.

Conditional qualitative treatment effects (CQTE)

- B and C are two disjoint subsets of $[1, \dots, p]$.
- X^B and X^C are the subsets of X .

Conditional qualitative treatment effects (CQTE)

- B and C are two disjoint subsets of $[1, \dots, p]$.
- X^B and X^C are the subsets of X .
- X^C have CQTE given X^B , if there exists some nonempty subsets \mathcal{C}_1 , \mathcal{C}_2 and \mathcal{B} such that
 - (i) $\Pr \{(X^B, X^C) \in \mathcal{B} \times \mathcal{C}_1\} > 0$ and $\Pr \{(X^B, X^C) \in \mathcal{B} \times \mathcal{C}_2\} > 0$.
 - (ii) For any $x_1^C \in \mathcal{C}_1$, $x_2^C \in \mathcal{C}_2$ and $x^B \in \mathcal{B}$, we have

$$\begin{aligned} & \arg \max_a E \left\{ Y^*(a) | X^B = x^B, X^C = x_1^C \right\} \\ \neq & \arg \max_a E \left\{ Y^*(a) | X^B = x^B, X^C = x_2^C \right\}. \end{aligned}$$

Unconditional qualitative treatment effects

- $B = \emptyset$.

Unconditional qualitative treatment effects

- $B = \emptyset$.
- $CQTE = QTE$.

Unconditional qualitative treatment effects

- $B = \emptyset$.
- CQTE = QTE.
- Let

$$\tau_0^C(x^C) = E\{\tau_0(X) | X^C = x^C\}.$$

No CQTE means $\tau_0^C(X^C) \geq 0$, a.s or $\tau_0^C(X^C) \leq 0$, a.s.

Unconditional qualitative treatment effects

- $B = \emptyset$.
- CQTE = QTE.
- Let

$$\tau_0^C(x^C) = E\{\tau_0(X) | X^C = x^C\}.$$

No CQTE means $\tau_0^C(X^C) \geq 0, a.s$ or $\tau_0^C(X^C) \leq 0, a.s$.

Conditional qualitative treatment effects

- B is not empty.

Unconditional qualitative treatment effects

- $B = \emptyset$.
- CQTE = QTE.
- Let

$$\tau_0^C(x^C) = E\{\tau_0(X) | X^C = x^C\}.$$

No CQTE means $\tau_0^C(X^C) \geq 0, a.s$ or $\tau_0^C(X^C) \leq 0, a.s$.

Conditional qualitative treatment effects

- B is not empty.
- Assume X^B have QTE.

Unconditional qualitative treatment effects

- $B = \emptyset$.
- CQTE = QTE.
- Let

$$\tau_0^C(x^C) = E\{\tau_0(X) | X^C = x^C\}.$$

No CQTE means $\tau_0^C(X^C) \geq 0, a.s$ or $\tau_0^C(X^C) \leq 0, a.s$.

Conditional qualitative treatment effects

- B is not empty.
- Assume X^B have QTE.
- CQTE measures whether X^C are “important” in decision making given X^B .

- For any $D \subseteq [1, \dots, p]$, define

$$\tau_0^D(x^D) = \mathbb{E}\{\tau_0(X) | X^D = x^D\},$$
$$d_{opt}^D(x) = I(\tau_0^D(x^D) > 0).$$

- For any $D \subseteq [1, \dots, p]$, define

$$\begin{aligned}\tau_0^D(x^D) &= \mathbb{E}\{\tau_0(X) | X^D = x^D\}, \\ d_{opt}^D(x) &= I(\tau_0^D(x^D) > 0).\end{aligned}$$

- Let $W = B \cup C$. Define

$$ER^{W,B} = \begin{cases} 0, & \text{if } \tau_0^W(X) = 0, \text{ a.s.} \\ \frac{\mathbb{E}[|d_{opt}^W(X) - d_{opt}^B(X)| I\{\tau_0^W(X^W) \neq 0\}]}{\Pr\{\tau_0^W(X^W) \neq 0\}}, & \text{otherwise,} \end{cases}$$

$$\begin{aligned}VD^{W,B} &= V(d_{opt}^W) - V(d_{opt}^B) \\ &= \mathbb{E}\left(\tau_0^W(X^W)[I\{\tau_0^W(X^W) \geq 0\} - I\{\tau_0^B(X^B) \geq 0\}]\right).\end{aligned}$$

Theorem (Charaterization of No CQTE)

Under certain conditions, the followings are equivalent:

- (i) X^c doesn't have CQTE given X^B .

Theorem (Charaterization of No CQTE)

Under certain conditions, the followings are equivalent:

- (i) X^c doesn't have CQTE given X^B .
- (ii) $VD^{W,B} = 0$.

Theorem (Charaterization of No CQTE)

Under certain conditions, the followings are equivalent:

- (i) X^c doesn't have CQTE given X^B .
- (ii) $VD^{W,B} = 0$.
- (iii) $ER^{W,B} = 0$.

Theorem (Charaterization of No CQTE)

Under certain conditions, the followings are equivalent:

- (i) X^c doesn't have CQTE given X^B .
- (ii) $VD^{W,B} = 0$.
- (iii) $ER^{W,B} = 0$.
- (iv) For any x^W such that $\tau_0^W(x^W) \neq 0$, $d_{opt}^W(x) = d_{opt}^B(x)$.

Theorem (Charaterization of No CQTE)

Under certain conditions, the followings are equivalent:

- (i) X^C doesn't have CQTE given X^B .
- (ii) $VD^{W,B} = 0$.
- (iii) $ER^{W,B} = 0$.
- (iv) For any x^W such that $\tau_0^W(x^W) \neq 0$, $d_{opt}^W(x) = d_{opt}^B(x)$.
- (v) For any fixed x^B , $\tau_0^W(x^B, x^C) \geq 0$ for any x^C or $\tau_0^W(x^B, x^C) \leq 0$ for any x^C .

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

- $\tau_0(x_1, x_2) = x_1 x_2^2$.

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

- $\tau_0(x_1, x_2) = x_1 x_2^2$.
- No CQTE.

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

- $\tau_0(x_1, x_2) = x_1 x_2^2$.
- No CQTE.
- $\tau_0(x_1, x_2) = x_1 \max(x_2, 0)$.

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

- $\tau_0(x_1, x_2) = x_1 x_2^2$.
- No CQTE.
- $\tau_0(x_1, x_2) = x_1 \max(x_2, 0)$.
- No CQTE.

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

- $\tau_0(x_1, x_2) = x_1 x_2^2$.
- No CQTE.
- $\tau_0(x_1, x_2) = x_1 \max(x_2, 0)$.
- No CQTE.
- $\tau_0(x_1, x_2) = x_1(x_2 - 2)$.

X_1, X_2 are independently generated from $\text{Unif}[-2, 2]$. Does X_2 has CQTE given X_1 ?

- $\tau_0(x_1, x_2) = x_1 x_2^2$.
- No CQTE.
- $\tau_0(x_1, x_2) = x_1 \max(x_2, 0)$.
- No CQTE.
- $\tau_0(x_1, x_2) = x_1(x_2 - 2)$.
- No CQTE.

- Hypothesis testing:

H_0 : X^C doesn't have CQTE given X^B ,

versus

H_1 : X^C has CQTE given X^B .

- Hypothesis testing:

H_0 : X^C doesn't have CQTE given X^B ,

versus

H_1 : X^C has CQTE given X^B .

- Such hypothesis assesses the incremental value of X^C in optimal treatment decision making conditional on X^B .

- Hypothesis testing:

H_0 : X^C doesn't have CQTE given X^B ,

versus

H_1 : X^C has CQTE given X^B .

- Such hypothesis assesses the incremental value of X^C in optimal treatment decision making conditional on X^B .
- If H_0 holds, estimate $d_{opt}^B(x)$.

- Hypothesis testing:

$$H_0 : X^C \text{ doesn't have CQTE given } X^B,$$

versus

$$H_1 : X^C \text{ has CQTE given } X^B.$$

- Such hypothesis assesses the incremental value of X^C in optimal treatment decision making conditional on X^B .
- If H_0 holds, estimate $d_{opt}^B(x)$.
- Otherwise, estimate $d_{opt}^W(x)$.

- Let f^W be the probability density function of X^W .

- Let f^W be the probability density function of X^W .
- Estimate $\tau_0^W(x^W)f^W(x^W)$ by

$$\tau_n^W(x^W) = \frac{1}{n} \sum_{i=1}^n \left(\frac{A_i}{\pi_i} - \frac{1 - A_i}{1 - \pi_i} \right) Y_i K_{h^W}^W(x^W - X_i^W),$$

for some multivariate kernel function $k_{h^W}^W$ with the bandwidth h^W .

Testing procedure

$$S_n^{W,B} = \int_{x^W \in \Omega^W} \tau_n^W(x^W) \{d_n^W(x^W) - d_n^B(x^B)\} I(x^W \notin \hat{E}) d\nu(x^W),$$

where

$$\hat{E} = \left\{ x^W : \left| \frac{\tau_n^W(x^W)}{\hat{f}^W(x^W)} \right| \leq \eta_n, \left| \frac{\tau_n^B(x^B)}{\hat{f}^B(x^B)} \right| \leq \eta_n \right\},$$

for some sequence $\eta_n \rightarrow 0$. Here, \hat{f}^W and \hat{f}^B are the kernel density estimators of f^W and f^B , respectively.

Testing procedure (Cont'd)

Let

$$\hat{F} = \{x^W : |\tau_n^W(x^W)/\hat{f}^W(x^W)| \leq \eta_n, |\tau_n^B(x^B)/\hat{f}^B(x^B)| > \eta_n\}.$$

The test statistic is defined by

$$T_n^{W,B} = \begin{cases} \{\sqrt{n}S_n^{W,B} - \hat{a}_n(\hat{F})\}/\hat{\sigma}_n(\hat{F}), & \text{if } \nu(\hat{F}) \neq 0, \\ \{\sqrt{n}S_n^{W,B} - \hat{a}_n(\Omega^W)\}/\hat{\sigma}_n(\Omega^W), & \text{otherwise.} \end{cases}$$

We reject the null when $T_n^{W,B} > z_\alpha$.

Theorem (Consistency)

Under certain conditions, when H_0 is true, we have

$$\lim Pr(T_n^{W,B} > z_\alpha) \leq \alpha,$$

for $0 < \alpha \leq 0.5$.

Theorem (Consistency)

Under certain conditions, when H_0 is true, we have

$$\lim \Pr(T_n^{W,B} > z_\alpha) \leq \alpha,$$

for $0 < \alpha \leq 0.5$. If $\Pr\{\tau_0^W(X^W) = 0\} = 0$, we have

$$\lim \Pr(T_n^{W,B} > z_\alpha) = 0.$$

Theorem (Consistency)

Under certain conditions, when H_0 is true, we have

$$\lim \Pr(T_n^{W,B} > z_\alpha) \leq \alpha,$$

for $0 < \alpha \leq 0.5$. If $\Pr\{\tau_0^W(X^W) = 0\} = 0$, we have

$$\lim \Pr(T_n^{W,B} > z_\alpha) = 0.$$

When H_1 is true, we have

$$\lim \Pr(T_n^{W,B} > z_\alpha) \rightarrow 1.$$

Theorem (Consistency)

Under certain conditions, when H_0 is true, we have

$$\lim Pr(T_n^{W,B} > z_\alpha) \leq \alpha,$$

for $0 < \alpha \leq 0.5$. If $Pr\{\tau_0^W(X^W) = 0\} = 0$, we have

$$\lim Pr(T_n^{W,B} > z_\alpha) = 0.$$

When H_1 is true, we have

$$\lim Pr(T_n^{W,B} > z_\alpha) \rightarrow 1.$$

Theorem (Informal statement)

Under certain conditions, $T_n^{W,B}$ has non-negligible powers against some nonstandard $n^{-1/2}$ local alternatives.

Simulation models:

$$Y = 1 - \frac{X_1 - X_2}{2} + A\phi_1(X_1)\phi_2(X_2) + e,$$

Table: Simulation results.

	<i>n</i>	VD = 0		VD = 4%		VD = 8%		VD = 12%	
		α level		α level		α level		α level	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Scenario 1	300	4.3%	6.0%	24.0%	34.0%	58.7%	68.1%	82.2%	87.5%
	600	1.5%	3.3%	36.7%	45.5%	75.8%	83.3%	95.7%	97.3%
Scenario 2	300	7.0%	11.1%	23.8%	32.7%	60.5%	69.3%	88.2%	92.3%
	600	5.5%	10.0%	31.0%	41.8%	83.0%	90.5%	98.3%	99.5%
Scenario 3	300	3.8%	6.5%	37.3%	48.5%	76.3%	79.7%	92.7%	94.7%
	600	2.7%	6.7%	52.5%	61.8%	99.2%	100%	99.8%	99.8%
Scenario 4	300	6.2%	9.8%	39.8%	47.7%	79.2%	87.3%	94.8%	96.7%
	600	5.2%	8.8%	59.3%	68.2%	96.8%	98.3%	99.5%	99.5%
Scenario 5	300	5.2%	9.7%	29.3%	40.5%	68.0%	76.3%	94.0%	96.8%
	600	5.3%	9.5%	36.7%	45.5%	75.8%	83.3%	95.7%	97.3%

- 1 Set $B = \emptyset$. In Step 1, for each variable i , define the set $W_i = \{i\}$ and calculate the p -value p_i for each test statistic $T^{W_i, B}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p -value in the set B , i.e.,

$$B \leftarrow \{\arg \min_i p_i\}.$$

- 1 Set $B = \emptyset$. In Step 1, for each variable i , define the set $W_i = \{i\}$ and calculate the p -value p_i for each test statistic $T^{W_i, B}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p -value in the set B , i.e,

$$B \leftarrow \{\arg \min_i p_i\}.$$

- 2 In Step 2, for each variable $i \notin B$, define $W_i = B \cup \{i\}$ and calculate the p -value p_i for each test statistic $T^{W_i, B}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p -value,

$$B \leftarrow B \cup \{\arg \min_i p_i\}.$$

- 1 Set $B = \emptyset$. In Step 1, for each variable i , define the set $W_i = \{i\}$ and calculate the p -value p_i for each test statistic $T^{W_i, B}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p -value in the set B , i.e.,

$$B \leftarrow \{\arg \min_i p_i\}.$$

- 2 In Step 2, for each variable $i \notin B$, define $W_i = B \cup \{i\}$ and calculate the p -value p_i for each test statistic $T^{W_i, B}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p -value,

$$B \leftarrow B \cup \{\arg \min_i p_i\}.$$

- 3 Continue the second step until it stops. Output B .

- Introduce the notion of CQTE and present several equivalent representations of No CQTE.

- Introduce the notion of CQTE and present several equivalent representations of No CQTE.
- Propose a testing procedure for testing No CQTE.

- Introduce the notion of CQTE and present several equivalent representations of No CQTE.
- Propose a testing procedure for testing No CQTE.
- Develop a procedure for selecting prescriptive variables in sequential order.

- Introduce the notion of CQTE and present several equivalent representations of No CQTE.
- Propose a testing procedure for testing No CQTE.
- Develop a procedure for selecting prescriptive variables in sequential order.
- Extend the testing procedure to a high-dimensional setting.

Thank you!