A Contemporaneous External Control Trial Design with Robust Causal Estimate of Rx Effect

Ming T. Tan, PhD, Professor
Dept. of Biostatistics, Bioinformatics & Biomathematics
Georgetown University
ming.tan@georgetown.edu

Presentation at Regulatory-Industry Stat 9/24/2020, Washington DC, USA

Work in collaboration with M. Atkins, S. Rapisuwon, A. Yuan and A. Yin.

Disclaimer: This presentation is a result of my methodology research. The views expressed belong solely to the author, and not necessarily to the author's employer, and sponsor.

Georgetown | Lombardi
COMPREHENSIVE CANCER CENTER

Outline

- Motivating example: ocular melanoma trial using a synthetic (RWE external) control
- Design and analysis challenges
- Robust causal inference using the enhanced doubly robust estimate of Rx effect
- Summary and future work

A motivational example of external control trial: Immunotherapy for Ocular Melanoma

- Adjuvant nivolumab combined with ipilimumab for adults with pretreated high-risk ocular melanoma, a rare disease
- Nivolumab 240 mg IV over 30 minutes given Day 1, 14 and 28 of each Cycle; Ipilimumab 1 mg/kg IV over 60 minutes given Day 1 of each Cycle
- RCT ruled out
- Open-label, single-arm, contemporaneous control, multi-center Phase II clinical trial
- Currently enrolling patients and registered June 11, 2018 in ClinicalTrials.gov with GU-LCCC (lead), Pitt, Dana-Farber Harvard, Yale, Northwestern and Colorado.

Contemporaneous controls

- External controls? Not much historical data available
- Ocular Melanoma Registry is a national collaborative registry to study the natural history and biology of OM, led by the Melanoma Research Foundation, and Patient Registry Working Group. It aims to enroll 200 new cases per year to the registry from major ocular oncology institutions
- Take controls from the ongoing registry but not participating in the trial
- This contemporaneous sample will have advantages over a true historical control as it will involve patients who would have been eligible for the trial and provide the opportunity to match risk criteria (age, gender, tumor site and thickness, somatic BAP1 status) in control and treated patients

Statistical Issues

- Primary endpoint: landmark 3-year Relapse Free Survival (RFS)
- Design: Survival at fixed points with censoring? How to determine sample size?
- Analysis: Though contemporaneous control, the trial is not randomized. Is the Rx effect causal? What is the most appropriate test?
- Monitoring: interim analysis for pronounced Rx effect or futility? Xiong, Tan, Boyett, SIM, 2007

Sample size determination

- Exact sample size determination for comparing correlated matched survival at a fixed time point is still under research.
- Estimated sample size is 50 with at least 80% power to detect the difference of 25% (50% vs 75%) in 3-year RFS.
- We adopted a hybrid approach combining study result of Su et al. 2014 and Costigan 2015.

Sample size

- Used a frailty model to capture the potential correlation between the pair. For a two-sided α and a given power $1-\beta$, the required sample size is

$$n = \frac{\sigma^2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\mu^2} \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$$

- Apply sample size calculation (R function) for different level of dependence θ we have

θ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
n	6	8	10	14	19	24	30	36	43	49

- Therefore, sample size of 50 patients per group will provide ≥80% power to show the difference of 25% in 3-year RFS.

Statistical Analysis Plan

- Randomization no longer applies
- How do we estimate the Rx effect Δ?
- Causal inference approach using the doubly robust estimates (Robins et al., JASA 1994)
- Enhanced doubly robust estimate (Yuan, Yin & Tan, 2019)

Causal Inference: Doubly Robust Estimation

- Let Y_{ij} be the potential outcome of subject *i* receiving treatment *j* (*j=1, 0*)
- The causal effect (Rubin 1974) is defined as $E(Y_{i1}) - E(Y_{i0})$

- But for each subject, either $Y_{i1}\ (\mathrm{t}_i = 1)$ or $Y_{i0}(\mathrm{t}_i = 0)$, i.e., only observed

$$Y_i = t_i Y_{i1} + (1 - t_i)\, Y_{i0}$$

- When the missing part is missing completely at random, or as in RCT, $E(Y_{i1}) - E(Y_{i0})$ can be estimated by the usual methods (e.g., difference of the sample means)

- In non-randomized setting, the estimate is biased and corrections are needed → causal inference e.g. doubly robust estimate approach which combines regression on outcome with a model for Rx assignment (i.e., the propensity score) to estimate the causal effect

Robbins et al. 1994

The propensity score model. The classical model for binary response with covariates is the logistic model

$$P(T=1|\boldsymbol{x}) = \frac{\exp(\boldsymbol{x}^T\gamma)}{1+\exp(\boldsymbol{x}^T\gamma)}.$$

For generality and model robustness we specify the model as

$$P(T=1|\boldsymbol{x}) = \pi(\boldsymbol{x}^T\boldsymbol{\gamma}), \quad \pi(\cdot)\in\Pi, \quad \|\boldsymbol{\gamma}\|=1, \qquad (1)$$

where Π is a collection of non-negative monotone increasing functions upper bounded by 1, and the constraint $\|\boldsymbol{\gamma}\|=1$ is for model identifiability. In particular, Π is the collection of all distribution functions. Note that when $dim(\boldsymbol{x})=1$, γ is absorbed into $\pi(\cdot)$ and there is no parameter $\boldsymbol{\gamma}$.

The regression model

In most literature about doubly robustness procedure, $m_j(\boldsymbol{x}, \boldsymbol{\beta}_j)$ are either specified parametric or by a known form of $m(\cdot)$, mostly the linear model. Here we specify the regression model as

$$y_{ij} = m_j(\boldsymbol{x}_i^T \boldsymbol{\beta}_j) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_j^2), \quad m_j(\cdot) \in \mathcal{M}, \quad j = 0, 1.$$

where $\boldsymbol{\beta}_j$ is the regression coefficients satisfying $||\boldsymbol{\beta}_j|| = 1$, $\mathcal{M}$ is the collection of bounded monotone increasing functions on R. The constraint $||\boldsymbol{\beta}_j|| = 1$ is for model identifiability. Note that when $dim(\boldsymbol{x}) = 1$, $beta = 1$ or there is no β in the model.

Estimate of Δ.

The treatment effect μ_j and Δ are jointly determined by $(\pi(\cdot), \boldsymbol{\gamma}, m_j(\cdot), \boldsymbol{\beta}_j : j = 0, 1)$. After obtaining $\hat{\boldsymbol{\gamma}}$ and $(\hat{\boldsymbol{\theta}}, \hat{m})$, the doubly robust estimate of Δ is given by

$$\hat{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{t_i y_i}{\hat{\pi}(\boldsymbol{x}_i^T\hat{\boldsymbol{\gamma}})} - \frac{(1-t_i)y_i}{1-\hat{\pi}(\boldsymbol{x}_i^T\hat{\boldsymbol{\gamma}})}$$

$$-\big(t_i - \hat{\pi}(\boldsymbol{x}_i^T\hat{\boldsymbol{\gamma}})\big)\Big(\frac{\hat{m}_1(\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_1)}{\hat{\pi}(\boldsymbol{x}_i^T\hat{\boldsymbol{\gamma}})} + \frac{\hat{m}_0(\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}_0)}{1-\hat{\pi}(\boldsymbol{x}_i^T\hat{\boldsymbol{\gamma}})}\Big). \qquad (4)$$

Rationale: Correct specification of either propensity π(.) or the regression model m_j(.) is challenging in practice. We have proposed robust models for both π(.) and m_j(.)

Asymptotic Properties

- Under proper regularity conditions, asymptotic normality of γ and β estimates still holds
- The proof is not straightforward (Yuan, Yin and Tan, 2020).

Test the significance

We are to test the null hypothesis $H_0 : \Delta = 0$ vs $H_1 : \Delta > 0$. The Wald test is convenient to use, under H_0,

$$T_n := \sqrt{n}\frac{\hat{\Delta}}{\sqrt{Var(\Delta_1)}} \xrightarrow{D} N(0,1).$$

If both the propensity score and regression models are correct, $Var(\Delta_1)$ can be estimated as, with $\tilde{\Delta}_i$ being the i-th summands in (4),

$$Var(\Delta_1) \approx \frac{1}{n}\sum_{i=1}^{n}(\Delta_i - \hat{\Delta})^2.$$

Simulation Studies

- For both propensity and response model, simulations are performed for n= 500, 1000, 1500. The treatment effects estimates from the proposed and four other methods are provided in the following table

Rx Effects Estimates from 5 Methods with Different Sample Sizes

	n=500			1000			1500		
			Δ			Δ			Δ
TRUE	18.903	15.123	3.781	18.903	15.123	3.781	18.903	15.123	3.781
Enhanced DB	19.058	15.252	3.806	18.974	15.221	3.753	18.938	15.196	3.741
(SD)	0.097	0.088	0.117	0.065	0.060	0.079	0.056	0.050	0.066
Parametric DB	17.563	15.474	2.089	17.515	14.790	2.725	17.371	16.392	0.979
(SD)	0.204	0.886	0.962	0.146	0.573	0.622	0.123	0.136	0.222
Parametric IPSW	14.456	16.790	-2.334	15.212	15.921	-0.709	16.029	15.785	0.245
(SD)	0.783	1.145	1.734	0.617	0.754	1.216	0.603	0.673	1.101
Semiparametric IPSW	16.125	15.495	0.630	15.406	14.112	1.294	16.599	13.822	2.777
SD	1.709	1.685	2.692	0.612	0.681	1.142	0.672	0.708	1.144
Naïve	19.199	14.893	4.306	19.206	14.896	4.310	19.203	14.896	4.307
(SD)	0.094	0.095	0.135	0.065	0.065	0.091	0.053	0.056	0.080

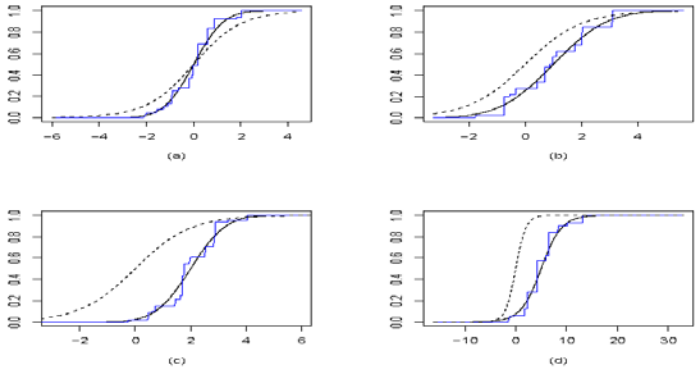

Figure 1: Propensity score model. True function (a): N(0, 4), (b): N(1, 1.5^2), (c): N(2,1), (d): Logistic(5, 2), solid black line), estimated function (solid step function), Logistic(0, 1, dotted line)

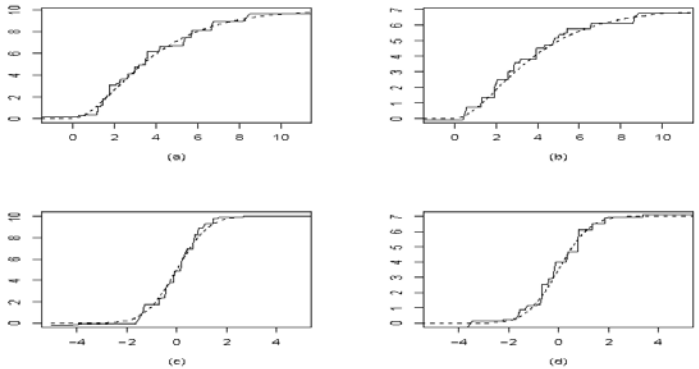

Figure 2: Regression model. (a). m_1: 10×Gamma(2, 2); (b). m_0: 7×Gamma(2, 2), (c). m_1: 10×N(0, 1); (d). m_0: 7×N(0,1). True link function in dotted line, estimated link function in solid line.

Back to the Trial: Survival Endpoint

- There is a scarcity of double robust methods. The existing one by Bai, Tsiatis, O'Brien (2013) is complex
- It involves three semiparametric quantities: propensity score, estimated survival function of censoring time, and survival function conditional on specific covariates.
- If two of the three are correctly specified, then the survival distribution is consistently estimated
- The enhanced DR method with semiparametric modeling involves only two quantities: the propensity score and conditional survival function

Conclusions and Ongoing Work

- We proposed an enhanced double robust estimator for causal inference
- The method is motivated by and will be used in analyzing the nonrandomizd “two- arm” treatment vs external control trial
- The improved statistical methods and the contemproraneous control from RWD provide a firmer basis to support causal inference on the treatment
- Details on surveil endpoints are being worked out