Causal inference on the difference of the restricted mean lifetime between two groups

work of P. Chen and A. Tsiatis (Biometrics 2001), among others

Tianchen Qian

Department of Biostatistics
Bloomberg School of Public Health
The Johns Hopkins University

SLAM Seminar, Mar 14, 2014

Outline of the Talk

The Data Problem

- **Data source:** observational study of acute coronary syndrome patients from Duke University Medical Center.
- **Duration of study:** 5 years. (start of 1994 - end of 1998)
- **Sample size:** 6033 patients. 3786 have been followed for 5+ years or died prior to the end of study (1998); the rest have censored survival times.
- **Treatment groups:** PCI group (3868 patients), MED group (2165 patients).[1]
- **Outcome of interest:** survival time up to 5 years.
- **Goal:** *Compare restricted mean lifetime between the two treatment groups, to assess treatment effect.*

[1] PCI: percutaneous coronary intervention. MED: medically treated.

Solution 1: compare group means directly

- Throw away censored data (assume non-informative censoring).
- Compare group means.
- **Cons:** loss efficiency.

Solution 2: use Kaplan-Meier estimate

- Denote survival function of group j as $S_j(t)$, $j = 0, 1$.
- Kaplan-Meier estimator $\hat{S}_j(t)$, using data from group j.
- Mean survival time:

$$\mu = E[T] = \int_0^L P(T \geq t)\, dt = \int_0^L S(t)\, dt, \tag{1}$$

  where $L = 5$ years.
- Difference between groups:

$$\hat{\delta} = \hat{\mu}_1 - \hat{\mu}_0 = \int_0^L \left\{ \hat{S}_1(t) - \hat{S}_0(t) \right\} dt. \tag{2}$$

- **Cons:** Not adjust for different covariate distribution between groups, so the estimated "treatment effect" is likely to be biased.

Solution 3: use Cox model for $\hat{S}_j(t)$

- Still use

$$\hat{\delta} = \hat{\mu}_1 - \hat{\mu}_0 = \int_0^L \left\{ \hat{S}_1(t) - \hat{S}_0(t) \right\} dt \tag{3}$$

  as the treatment effect estimator.
- Estimate $\hat{S}_j(t)$ using Cox's proportional hazards model, which can incorporate covariate information in the model.
- **This is the model we focus on.**

Notations

- T_i: restricted survival time ($\leq L$).
- C_i: censoring time.
- $\Delta_i = I(T_i \leq C_i)$: censoring indicator.
- $X_i = \min(T_i, C_i)$: observed failure time.
- Z_i: covariate vector.
- $N_i(t) = I(X_i \leq t, \Delta_i = 1)$.
- $Y_i(t) = I(X_i \geq t)$.
- $M_i(t) = N_i(t) - \int_0^t \lambda_i(u) Y_i(u) du$.
- $M(t) = \sum_{i=1}^n M_i(t)$, $N(t) = \sum_{i=1}^n N_i(t)$, $Y(t) = \sum_{i=1}^n Y_i(t)$.

Review of Cox model

- Assume

$$\lambda(t \mid Z) = \lambda_0(t)\, e^{\beta^T Z}, \tag{4}$$

where $\lambda_0(t)$ is the unspecified baseline hazard.

- The estimator $\hat{\beta}$ is the maximizer of the *partial likelihood function*:

$$L_P(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta^T Z_i(x_i)}}{\sum_{j \in \mathscr{R}_i} e^{\beta^T Z_j(x_i)}} \right\}^{\delta_i}, \tag{5}$$

where $x_1, \ldots, x_n$ are n observed survival times. $\mathscr{R}_i = \{j : x_j \geq x_i\}$ is the risk set, and $\delta_i = I(t_i \leq c_i)$ is the observed version of Δ_i.

Review of Cox model (continued)

- We will use Breslow's estimator (Breslow, 1972 JRSSB [2]) to estimate the cumulative baseline hazard:

$$\hat{\Lambda}_0(t) = \sum_{x_i \leq t} \frac{\delta_i}{\sum_{j \in \mathscr{R}_i} e^{\hat{\beta}^T Z_j(x_i)}}. \tag{6}$$

- With the above definitions, Breslow's estimator can be rewritten as:

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u) e^{\hat{\beta}^T Z_i}}. \tag{7}$$

- Asymptotic results: Andersen and Gill, 1982 Annals of Statistics[1].
- Basic idea: use counting process martingale representation, then apply martingale central limit theorem. See Fleming and Harrington's book "Counting Process and Survival Analysis" [4] for a good reference.

Rubin's causal model (very brief)

- For individual i, define T_i^0 and T_i^1 to be the outcome if the individual were assigned treatment 0 or 1.
- Individual causal treatment effect: $\delta_i = T_i^1 - T_i^0$.
- Average causal treatment effect for a group of people:

$$\delta = \frac{1}{n}\sum_{i=1}^{n}\delta_i = \left(\frac{1}{n}\sum_{i=1}^{n}T_i^1\right) - \left(\frac{1}{n}\sum_{i=1}^{n}T_i^0\right). \tag{8}$$

This can be estimated by

$$\hat{\delta} = \left(\frac{1}{n}\sum_{i=1}^{n}\hat{T}_i^1\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\hat{T}_i^0\right). \tag{9}$$

Our estimator

- According to Rubin's model, we want to compare:
  - the restricted mean lifetime if everyone were in treatment group 1.
  - the restricted mean lifetime if everyone were in treatment group 0.
- So the estimator is:

$$\hat{\delta} = \int_0^L \left\{ \hat{S}_1(u) - \hat{S}_0(u) \right\} du \tag{10}$$

$$= \int_0^L \left\{ \frac{1}{n} \sum_{i=1}^{n} \hat{S}_1(u \mid Z_i) - \frac{1}{n} \sum_{i=1}^{n} \hat{S}_0(u \mid Z_i) \right\} du. \tag{11}$$

- We estimate the above using two different Cox models.

Two models

- Consider two models. (A being the treament indicator.)
- Model 1:

$$\lambda\left(t \mid A=0, Z\right) = \lambda_0\left(t\right) e^{\beta_0^T Z}, \tag{12}$$

$$\lambda\left(t \mid A=1, Z\right) = \lambda_1\left(t\right) e^{\beta_1^T Z}. \tag{13}$$

- Model 2:

$$\lambda\left(t \mid A, Z\right) = \lambda_0\left(t\right) e^{\gamma_0 A + \gamma_1^T Z} = \lambda_0\left(t\right) e^{\gamma^T W}, \tag{14}$$

  where $\gamma = \left(\gamma_0, \gamma_1^T\right)^T$ and $W = \left(A, Z^T\right)^T$.
- Bias-variance tradeoff.

Estimate parameters in model 1

- For model 1,

$$\lambda\left(t \mid A=0, Z\right) = \lambda_0\left(t\right) e^{\beta_0^T Z}, \tag{15}$$

$$\lambda\left(t \mid A=1, Z\right) = \lambda_1\left(t\right) e^{\beta_1^T Z}. \tag{16}$$

  - Use individuals in treatment group 0 to estimate $\hat{\beta}_0$ and $\hat{\Lambda}_0\left(u\right)$:

$$\hat{\Lambda}_0\left(u\right) = \int_0^u \frac{\sum_{i=1}^n \left(1-A_i\right) dN_i\left(t\right)}{\sum_{i=1}^n \left(1-A_i\right) e^{\hat{\beta}_0^T Z_i} Y_i\left(t\right)}. \tag{17}$$

  - Use individuals in treatment group 1 to estimate $\hat{\beta}_1$ and $\hat{\Lambda}_1\left(u\right)$.
  - $\hat{S}_j\left(u \mid Z_i\right) = \exp\left\{-\hat{\Lambda}_j\left(u\right) \exp\left(\hat{\beta}_j^T Z_i\right)\right\}$, $j = 0, 1$.
  - $\hat{\delta} = \int_0^L \left\{\frac{1}{n}\sum_{i=1}^n \hat{S}_1\left(u \mid Z_i\right) - \frac{1}{n}\sum_{i=1}^n \hat{S}_0\left(u \mid Z_i\right)\right\} du$.

Estimate parameters in model 2

- For model 2,

$$\lambda\left(t \mid A, Z\right) = \lambda_0\left(t\right) e^{\gamma_0 A + \gamma_1^T Z} = \lambda_0\left(t\right) e^{\gamma^T W}, \tag{18}$$

where $\gamma = \left(\gamma_0, \gamma_1^T\right)^T$ and $W = \left(A, Z^T\right)^T$.

  - Use all the data from both treatment groups to get $\hat{\gamma}$ and $\hat{\Lambda}_0\left(u\right)$:

$$\hat{\Lambda}_0\left(u\right) = \int_0^u \frac{\sum_{i=1}^n dN_i\left(t\right)}{\sum_{i=1}^n e^{\hat{\gamma}^T W_i} Y_i\left(t\right)}. \tag{19}$$

  - $\hat{S}_0\left(u \mid Z_i\right) = \exp\left\{-\hat{\Lambda}_0\left(u\right) \exp\left(\hat{\gamma}_1^T Z_i\right)\right\}$,
    $\hat{S}_1\left(u \mid Z_i\right) = \exp\left\{-\hat{\Lambda}_0\left(u\right) \exp\left(\hat{\gamma}_0 + \hat{\gamma}_1^T Z_i\right)\right\}$.
  - $\hat{\delta} = \int_0^L \left\{\frac{1}{n}\sum_{i=1}^n \hat{S}_1\left(u \mid Z_i\right) - \frac{1}{n}\sum_{i=1}^n \hat{S}_0\left(u \mid Z_i\right)\right\} du$.

Var $\left(\hat{\delta}\right)$: Influence function

Definition

Let $\mathbf{X}_n = (X_1, \ldots, X_n)$, with X_i i.i.d. following some probability model. Suppose we are interested in estimating some parameter γ, whose true value is γ_0. An estimator $\hat{\gamma}(\mathbf{X}_n)$ of γ is said to be ***asymptotically linear***, if there exists $\varphi(x)$, such that

$$\sqrt{n}\left(\hat{\gamma}(\mathbf{X}_n) - \gamma_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i) + o_P(1), \tag{20}$$

with $E[\varphi(X)] = 0$ and $E\left[\varphi(X)\varphi(X)^T\right]$ finite and non-singular. The function $\varphi(x)$ is called the ***influence function*** for the estimator $\hat{\gamma}(\mathbf{X}_n)$.

- Useful in computing asymptotic variance.

Var $\left(\hat{\delta}\right)$: derive IF of $\hat{\delta}$

- General idea:
  1. Derive influence functions for $\hat{S}_0(u)$ and $\hat{S}_1(u)$. Use Andersen and Gill's result (1982).
  2. Derive influence functions for $\int_0^L \hat{S}_0(u)\,du$ and $\int_0^L \hat{S}_1(u)\,du$.
  3. Derive influence functions for $\hat{\delta} = \int_0^L \hat{S}_1(u)\,du - \int_0^L \hat{S}_0(u)\,du$.

Simulation 1

- Under strong null hypothesis: $H_0^* : S_1(u \mid Z) = S_0(u \mid Z)$ for all Z.
  - $Z \sim N(0,1)$.
  - $P(A = 1 \mid Z) = e^Z / (1 + e^Z)$.
  - $T^0, T^1 \sim$ Exponential $\left(e^{1+4Z}\right)$.
  - Independent censoring: $C \sim$ Exponential (0.1).
  - $L = 12.04$.

	Strong Null Hypothesis $\delta = 0$		
	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_{KM}$
Bias	.0289	.0019	-3.0417
se $\left(\hat{\delta}\right)$	.2297	.1124	.5114
$\widehat{\text{se}}\left(\hat{\delta}\right)$	.2302	.1125	.5136
Coverage Prob.	.9470	.9520	.0000

[2]

[2] Table is extracted from Chen and Tsiatis, 2001 [3].

Simulation 2

- Under Alternative hypothesis: (Here Model 2 is misspecified.)
  - $Z \sim N(0,1)$.
  - $P(A=1 \mid Z) = e^Z / (1+e^Z)$.
  - $T^0 \sim \text{Exponential}\left(e^{1+4Z}\right)$, $T^1 \sim \text{Exponential}\left(e^{-2+3Z}\right)$.
  - Independent censoring: $C \sim \text{Exponential}(0.1)$.
  - $L = 12.04$.

	Alternative Hypothesis $\delta = 3.0662$		
	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_{KM}$
Bias	.0085	.7960	-3.3024
se $\left(\hat{\delta}\right)$	.2744	.2047	.5479
$\widehat{\text{se}}\left(\hat{\delta}\right)$	.2780	.2276	.5475
Coverage Prob.	.9524	.0404	.0000

[3]

[3] Table is extracted from Chen and Tsiatis, 2001 [3].

Simulation 3

- When no confounding.
  - $Z_1, Z_2, Z_3 \sim N(0,1)$.
  - $P(A=1 \mid Z_2) = e^{5Z_2}/\left(1+e^{5Z_2}\right)$.
  - $T^0 \sim$ Exponential $\left(e^{1+4Z_1}\right)$, $T^1 \sim$ Exponential $\left(e^{-2+3Z_1}\right)$.
  - Independent censoring: $C \sim$ Exponential (0.1).
  - $L = 12.04$.

		Bias	SSE	SEE	CP
	Z_1	-.0055	.2807	.2833	.9500
	Z_2	-.0666	.7692	.7762	.9426
	Z_3	-.0294	.5224	.5311	.9536
$\hat{\delta}_1$	Z_1, Z_2	-.0081	.3369	.3378	.9518
	Z_1, Z_3	-.0055	.2822	.2834	.9504
	Z_2, Z_3	-.0660	.7711	.7759	.9414
	Z_1, Z_2, Z_3	-.0076	.3390	.3382	.9518
$\hat{\delta}_{KM}$		-.0207	.5240	.5311	.9538

[4]Table is extracted from Chen and Tsiatis, 2001 [3].

Data example

- The data mentioned at the beginning of the talk.
- **Data source:** Observational study of acute coronary syndrome patients from Duke University Medical Center.
- **Duration of study:** 5 years. (start of 1994 - end of 1998)
- **Sample size:** 6033 patients. 3786 have been followed for 5+ years or died prior to the end of study (1998); the rest have censored survival times.
- **Treatment groups:** PCI group (3868 patients), MED group (2165 patients).[5]
- **Goal:** *Compare restricted mean lifetime between the two treatment groups, to assess treatment effect.*

[5] PCI: percutaneous coronary intervention. MED: medically treated.

Data example: Result

	$\hat{\delta}_1$	$\hat{\delta}_2$	$\hat{\delta}_{KM}$
Estimate	.1760	.1725	.3621
Standard error	.0377	.0355	.0419

The authors have also carried out a careful examination of the distribution of covariates by treatment (not presented in the article), suggesting that patients assigned medication are prognostically worse on average. Thus, one would expect that adjusting for prognostic factors would result in a smaller treatment difference.

Summary

- Used Cox model to compare restricted mean lifetime between two treatment groups.
- Constructed estimators and obtained their asymptotic distribution.
- Showed bias-variance tradeoff comparing two Cox models.

Related work

- Zhao et al. (Zhao et al., 2013 JASA[6]) used Cox model to estimate treatment effect in survival time between groups, and identified which subgroup benefits the most from the treatment (i.e. among which subgroup of people the treatment effect is the largest).
- Hubbard et al.'s approach (Hubbard, van der Laan and Robins, 2000[5]) for estimating the average causal treatment difference in survival in observational studies is through the use of inverse probability weighted estimators. They modeled both the censoring distribution and the propensity score as functions of the covariates.

Reference

P. Andersen and R. Gill.

Cox's regression model for counting processes: a large sample study.
The annals of statistics, pages 1100–1120, 1982.

N. E. Breslow.

Contribution to the discussion on the paper by d. r. cox, regression models and life tables.
JR stat soc B, 34(2):216–217, 1972.

P. Chen and A. Tsiatis.

Causal inference on the difference of the restricted mean lifetime between two groups.
Biometrics, 57(4):1030–1038, 2001.

T. Fleming and D. Harrington.

Counting processes and survival analysis, volume 169.
John Wiley & Sons, 2011.

A. Hubbard, M. Van Der Laan, and J. Robins.

Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies.
pages 135–177, 2000.

Lihui Zhao, Lu Tian, Tianxi Cai, Brian Claggett, and Lee-Jen Wei.

Effectively selecting a target population for a future comparative study.
Journal of the American Statistical Association, 108(502):527–539, 2013.

Asymptotic property of $\hat{\beta}$ and $\hat{\Lambda}(t)$

Define

$$S^{(k)}(\beta, t) = \frac{1}{n}\sum_{i=1}^{n} Z_i^{\otimes k} Y_i(t) e^{\beta^T Z_i}, \qquad k = 0, 1, 2, \tag{21}$$

where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$, $a^{\otimes 2} = aa^T$. Also, define

$$E(\beta, t) = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)}, \tag{22}$$

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - E(\beta, t)^{\otimes 2}, \tag{23}$$

$$s^{(k)}(\beta, t) = E\left[Z^{\otimes k} Y(t) e^{\beta^T Z}\right], \qquad k = 0, 1, 2, \tag{24}$$

$$e(\beta, t) = \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)}, \tag{25}$$

$$v(\beta, t) = \frac{s^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} - e^{\otimes 2}(\beta, t), \tag{26}$$

and the matrix

$$\Sigma = \int_0^L v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt. \tag{27}$$

Then under regularity conditions on the covariates and the amount of censoring, we have

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \xrightarrow{D} N\left(0, \Sigma^{-1}\right). \tag{28}$$

Asymptotic property of $\hat{\beta}$ and $\hat{\Lambda}(t)$

And under regularity conditions, $\sqrt{n}\left(\hat{\Lambda}-\Lambda_0\right)$ converges weakly on $D[0,L]$ (Space of Cadlag functions, equipped with Skorohod metric) to a Gaussian process with zero mean, independent increments, and variance function

$$\int_0^t \frac{\lambda_0(x)}{s^{(0)}(\beta_0,x)}dx + Q(\beta_0,t)^T \Sigma^{-1} Q(\beta_0,t), \tag{29}$$

where the vector function Q is given by

$$Q(\beta_0,t) = \int_0^t e(\beta_0,x)\lambda_0(x)\,dx. \tag{30}$$

Influence function for $\hat{\delta}_1$

$$\int_0^L \left[g_1^T \left\{ Z_i - \mu_1 \left(t, \beta_1 \right) \right\} - \frac{h_1 \left(t \right)}{s_1^{(0)} \left(t, \beta_1 \right)} \right] A_i dM_i \left(t \right) \tag{31}$$

$$- \int_0^L \left[g_0^T \left\{ Z_i - \mu_0 \left(t, \beta_0 \right) \right\} - \frac{h_0 \left(t \right)}{s_0^{(0)} \left(t, \beta_0 \right)} \right] \left(1 - A_i \right) dM_i \left(t \right) \tag{32}$$

$$+ \int_0^L \left[\left\{ S_1 \left(u \mid Z_i \right) - S_0 \left(u \mid Z_i \right) \right\} - \left\{ S_1 \left(u \right) - S_0 \left(u \right) \right\} \right] du, \tag{33}$$

where

$$g_j = \int_0^L b_j \left(u \right) du, \tag{34}$$

$$h_j \left(t \right) = \int_t^L c_j \left(u \right) du, \tag{35}$$

$$c_j \left(u \right) = E \left\{ S_j \left(u \mid Z \right) e^{\beta_0^T Z} \right\}, \ \pi = P \left\{ A = 0 \right\}, \tag{36}$$

$$b_j \left(u \right) = \left(\pi \Sigma_j \right)^{-1} \int_0^u \left[\mu_j \left(t, \beta_j \right) E \left\{ S_j \left(u \mid Z \right) e^{\beta_j^T Z} \right\} - E \left\{ Z S_j \left(u \mid Z \right) e^{\beta_j^T Z} \right\} \right] \lambda_j \left(t \right) dt, \tag{37}$$

$$s_j^{(0)} \left(t, \beta_j \right) = E \left[Y \left(t \right) e^{\beta_j^T Z} \right], \quad j = 0, 1. \tag{38}$$