Propensity Scores

By: Mei-Jie Zhang, Ph.D.

Medical College of Wisconsin, Division of Biostatistics

Friday, March 29, 2013

12:00-1:00 pm

The Medical College of Wisconsin is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

The Medical College of Wisconsin designates this live activity for a maximum of 1.0 *AMA PRA Category 1 Credit™*. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Hours of Participation for Allied Health Professionals

The Medical College of Wisconsin designates this activity for up to 1.0 hours of participation for continuing education for allied health professionals.

Financial Disclosure

- In accordance with the ACCME® standard for Commercial Support Number 6, all in control of content disclosed any relevant financial relationships. The following in control of content had **no** relevant financial relationships to disclose.

Name:	Role in Meeting:
Kwang Woo Ahn, PhD	Activity Director
Haley Montsma, BBA	Planning Committee
Mei-Jie Zhang, PhD	Presenter

Evaluation Forms

Your opinion matters!

Help us plan future meetings, by completing and submitting your evaluation forms.

Thank you.

Outlines

- Background/Motivation
- What is propensity score?
- Applications of propensity score
- How to analyze treatment effect using propensity score?
- Propensity score matching
- How to analyze matched pair data?
- Example of BMT study using propensity score
- Conclusions

Motivation

- **In medical studies physicians and patients often want to know the treatment effect:**
- **Example:** A cohort study comparing survival outcome of diffuse large B cell lymphoma patients underwent HLA-identical sibling allogeneic transplantation versus autologous transplantation (Lazarus et al., *BBMT*, 2010)
- **This is a retrospective cohort study:**
- Auto HCT often uses PB (peripheral blood) and
- HLA-identical sibling HCT often uses BM (bone marrow) and for high risk patients;

  Allo HCT patient needs to have a HLA-matched sibling donor available
- This **treatment selection bias** will introduce conclusion bias:

  ➔ **Estimate treatment effect will be biased**
- **Without adjustment we may comparing apples to oranges.**

Motivation

- **Randomized clinical trial is the gold standard method for analyzing treatment effect**
- **BUT, Randomized clinical trial cannot be done for all studies:**
  - Infeasible or unethical to assign patients to different treatment
  - In stem cell transplant study, not all patients are having HLA-identical sibling donor available
  - Cost and time related issues

- **Researches need to analyze treatment effect for observational cohort study using data from non-randomized clinical trial:**
  - want to show the difference in outcome is attributable to difference in treatment, not due to patient selection bias

Purpose of Propensity Score

- **Propensity score approach has been proposed to mimic the clinical trial**
- Propensity score approach can be used to **produce unbiased comparison of treatment effect** under some nonrandomized conditions
- Propensity score approach has been wildly adapted recently:

-- Publications in Pub Med with phrase of “propensity score”:

1983 – 1997: 0 – 10/year

1998 – 2002: 11 – 50/year

2003 – 2004: 51 – 100/year

2005 – 2006: 150 – 170/year

2007: 250/year

Treatment Causal Effect

- $R^{(A)}$: response if subject receives treatment

  A=1 (case: treated) or 0 (control)

- The **TRUE treatment causal effect** is:

$$E\{R^{(1)}\} - E\{R^{(0)}\} \qquad (1)$$

- But, the "observed" expected difference in response is

$$E\{R^{(1)} | A=1\} - E\{R^{(0)} | A=0\} \qquad (2)$$

  for a randomly selected treated subject (A=1) compare to a randomly selected control subject (A=0)

- For an observational study,

$$(1) \neq (2)$$

  since subjects are not randomly assigned to treatments

Randomized Clinical Trial

- In a randomized clinical trial, treatment assignment ***A*** and response $(R^{(1)}, R^{(0)})$ are conditionally **independent** give covariates (risk factors) ***Z***
- Under this conditional independent assumption, one can show that

(1) = (2)

- **Randomized clinical trial leads to an unbiased estimate for the treatment causal effect**

What is Propensity Score?

- **Propensity score is the probability of receiving treatment (*A*=1) given covariates *Z*,**

$$\pi(Z) = P(A=1 \mid Z)$$

Assuming $0 < \pi(Z) < 1$

- Note: **There is a positive probability, $\pi(Z) > 0$, even he/she received a ``NO" treatment (*A*=0)**.
- Rosenbaum and Robin (1983, *Biometrika*) showed that **conditioning on the propensity score allows for unbiased estimation of treatment effect**
- **Balancing propensity score mimics a randomized clinical trial**

Propensity Score Estimation

- **True propensity score is unknown: It needs to be estimated**
- **Logistic regression modeling is typically used to estimate the unknown propensity score:** based on a set of covariates which determent the treatment assignment
- Propensity score ranges from 0 to 1
- Propensity score estimation only depends on known covariates (such as patient characteristics), **it is independent of OUTCOMES**

Use of Propensity Score

- Rosenbaum and Rubin suggested **three common methods to estimate treatment effect using propensity score:**

1. **Stratification (sub-classification)**
2. **Matching**
3. **Covariate adjustment (treating propensity score as a covariate)**

Propensity Score Matching

- **The most common implementation of propensity score matching is pair matching, where pairs of treated (A=1) and untreated (A=0) subjects with similar propensity scores are formed**
- Similar propensity score is often defined as within a constant of standard deviation of the propensity score
- For example, 0.5 or 1 of standard deviation. Then,

1. Randomly select one matched control among all possible matched controls *OR*
2. Select one matched control with smallest difference in propensity score

How to Analyze Propensity Score Matched Pair Data

- To estimate the treatment effect, propensity score matching is often used **to reduce or eliminate** the effect of treatment selection bias
- A systematic **review of medical literature** between 1996 and 2003 by Austin (*Int. J. of Biostat*., 2009) found that **majority of studies did not account for the matched nature**
- Similar findings were observed in cardiovascular surgery literature and general cardiology literature

How to Analyze Propensity Score Matched Pair Data

- **Question: How to analyze matched pair data?**

  Do we need to count for the matching?
- Matched subjects are, on average, more similar in baseline covariates than two randomly selected treated and untreated subjects

  ➔ **propensity score matched sample are not from independent observations**
- Austin's simulation study showed that **adjusted matched-pair analysis will be needed to have correct nominal level of type I error**

HTC Study Using Propensity Score Matched-Pair Analysis

- For illustration purpose, we consider a **CIBMTR** (Center for International Blood and Marrow Transplant Research) study
- The CIBMTR is comprised of clinical and basic scientists share data on their blood and bone marrow transplant patients with CIBMTR Data Collection Center and Statistical Center located at the **Medical College of Wisconsin**
- The CIBMTR is a repository of information about results of transplants at more than **500 transplant centers worldwide**

HCT Study

- CIBMTR study compared outcomes of HLA-identical sibling allogeneic versus autologous transplantation for **diffuse large B-cell lymphoma (DLBCL)**
- 916 adults DLBCL patients (age: 18-60) underwent **1st Auto HCT (N=837)** or **myeloablative (MA) HLA-id sibling Allo HCT (N=79)** between 1995 and 2003
- Patients were reported to the CIBMTR by 156 centers in 17 different countries
- **This is a retrospective study (not a randomized clinical trial study)**

HCT Study

- **Key risk factors are imbalanced between cohorts:**

	Auto	Allo	P
Disease Stage III/IV	27%/39%	8%/62%	0.003
B Symptoms at DX	46%	58%	0.04
Ext Dis Involvement	57%	70%	0.02
Marrow Involvement	17%	42%	<0.001
Chemo Resistant Disease	15%	42%	<0.001
Graft Score: BM	9%	37%	<0.001
DX to TX, median, M	13(2-287)	11(2-156)	0.03

➔ **Allogeneic HCT for high risk patients**

HCT Study

- **ALLO HCT often uses BM for HIGH RISK patients**
- **Risk factors are imbalanced:**

  ➔ Directly estimate treatment effect will be biased

- **CIBMTR study performed a matched pair comparison** of the ALLO HCT group with a subset of closely matched AUTO HCT patients selected based on propensity score matching

HCT Study

- **Propensity score was calculated based on a fitting a logistic-regression model** with key risk factors of:

AGE	*SEX*
Karnofsky performance score	*B symptoms at DX*
Disease stage at diagnosis	*Extranodal disease at DX*
Marrow Involvement at DX	*# of prior-chemo*
Sensitivity to chemo	*Time from DX to TX*
Graft source	*Year of TX*

- **Propensity Score: Median (range; SD)**

-- **Combined Sample (N=893):** 0.042 (0.002 – 0.895; 0.123)

-- **Case (ALLO: N=79):** 0.236 (0.008 – 0.895; 0.239)

-- **Control (AUTO: N=814):** 0.039 (0.002 – 0.624; 0.086)

HCT Study

- **Propensity Score (PS) Matching:**

1. For each ALLO HCT patient (Case). Any AUTO HCT patient (Control) with a difference in propensity score of less than SD=0.123 was considered as a potential matched control. A matched control with smallest difference in PS was selected
2. Matching procedure (Step 1) was performed for each case
3. Matching Step 1 and 2 were repeated 4 times for possible 1 to 4 matching

- **Matching Result:**

-- **1 – 4 matching = 49 pairs; 1 – 3 matching = 2 pairs**

-- **2 – 3 matching = 12 pairs; 1 – 1 matching = 6 pairs**

-- 10 case cannot find any matched control

HCT Study

- **Analyzing matched time-to-event data:**
- **Variable:**

  allo=1 for ALLO HCT; =0 for AUTO HCT

  intxsurv=time from TX to death or end of FU (month)

  dead=1 for died; =0 for alive

  ipair: index of matched pair

1. Stratified Cox model (SAS Code):
   proc phreg; model intxsurv*dead(0)=allo; strata ipair;
2. Marginal Cox model (SAS Code):

   proc phreg covs(aggregate);

   model intxsurv*dead(0)=allo; id=ipair;

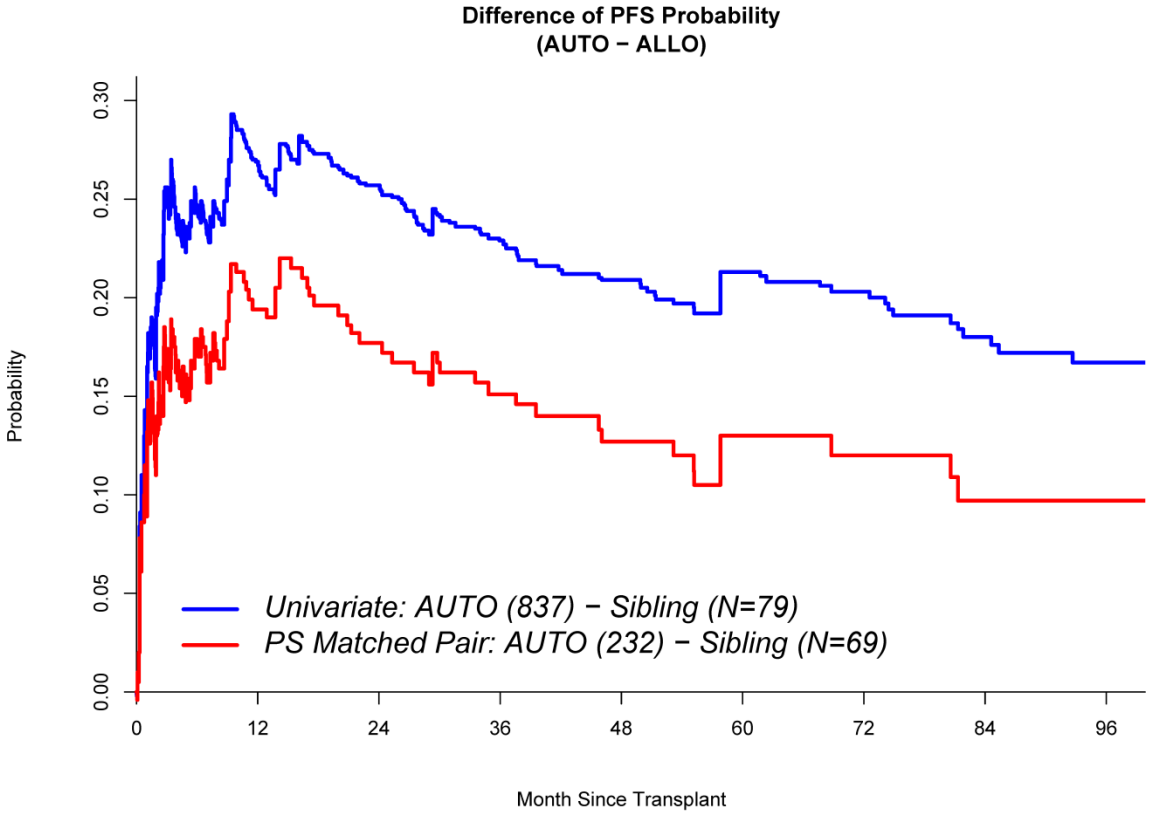
HCT Study

- **Summary of outcomes from matched pair comparison:**

	RR (95% CI)	P
Treatment Failure (1 – PFS)		
Allo vs Auto (overall)	1.95 (1.34 – 2.83)	0.0005
Within first 2 months after HCT	2.04 (1.38 – 3.01)	0.0003
Beyond first 2 months after HCT	1.19 (0.32 – 4.40)	0.7948
Mortality (1 – Survival)		
Allo vs Auto (overall)	2.38 (1.68 – 3.53)	<0.0001
Within first 2 months after HCT	2.77 (1.81 – 4.25)	<0.0001
Beyond first 2 months after HCT	1.05 (0.38 – 2.93)	0.9232

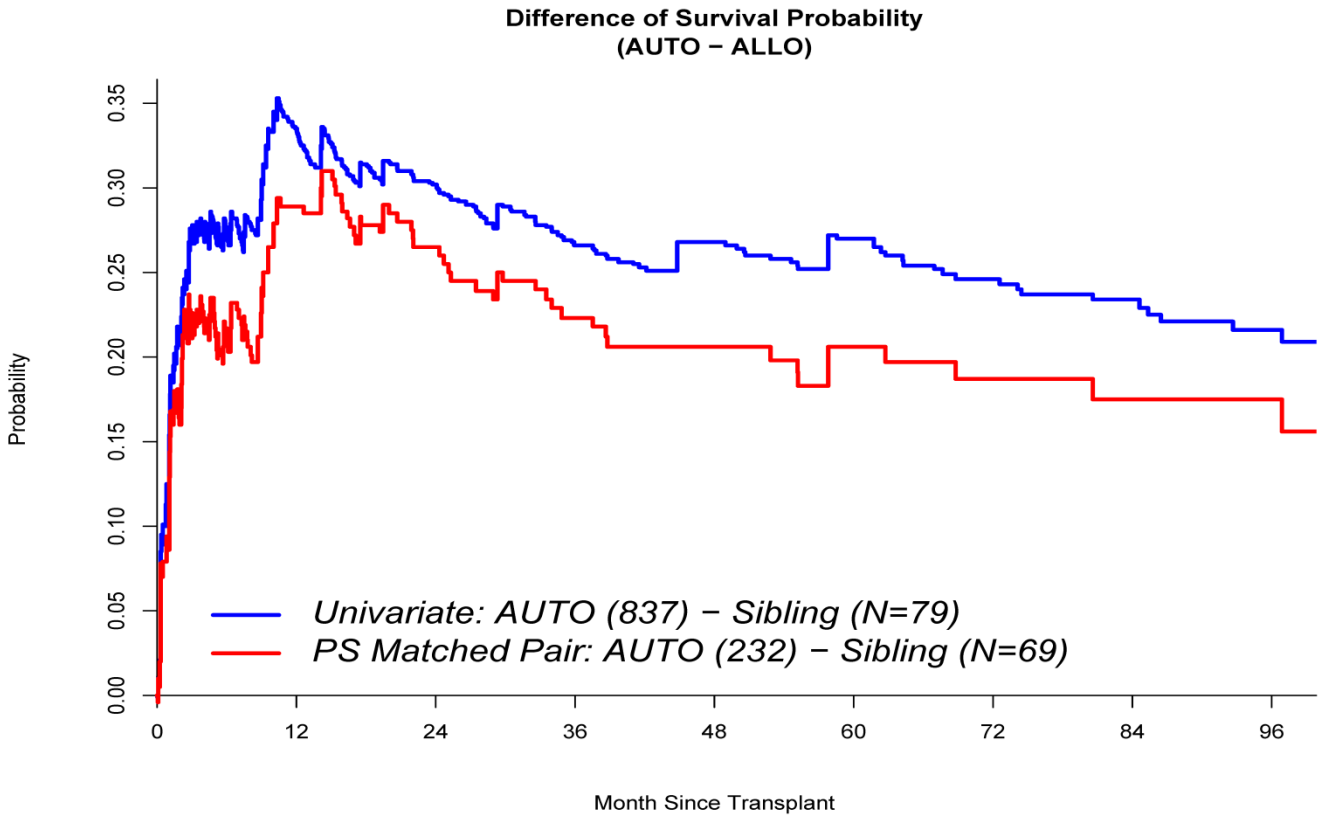
HCT Study

- Over Estimate AUTO Treatment Effect without Adjusting the Imbalance of Key Risk Factors between Cohorts:



HCT Study

- Over Estimate AUTO Treatment Effect without Adjusting the Imbalance of Key Risk Factors between Cohorts:



Conclusions

- **Benefits:**
  - Useful and unbiased method to analyzing treatment effect
  - Easy to adjusting a large number of risk factors
  - Useful matched design for saving time and money
  - Independent of outcomes
- **Limitations:**
  - Only adjust observed risk factors (Randomized clinical trial balance all risk factors (observed & unobserved))
  - Bias still may occur (To obtain unbiased inference using propensity score needs some assumptions/conditions. These assumptions/conditions may not hold)

References

- Austin, P.C. Type I error rates, converge of confidence intervals, and variance estimation in propensity score matched analysis. *The Inter. J. Biostat*, **5.1** (Article 13), 2009
- D'Agostino, R.B. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat In Med*, **17**, 2265-2281, 1998.
- Lazarus, H.M, Zhang, M.J. et al. A comparison of HLA-identical sibling allogeneic versus autologous transplantation for diffuse large B cell lymphoma. *BBMT*, **16**, 35-45, 2010
- Rosenbaum, P.R. and Rubin, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55, 1983.

Questions?