Part 3: Methods for Propensity Analysis

SMDM Short Course on Propensity

Charles Thomas

Center for Health Care Research & Policy

CWRU

Outline

- Definition of the Propensity Score
- Estimating the Propensity Score
- Using Propensity Scores
  - In Matching
  - For Subclassification
  - In Multivariate Adjustment
  - Subgroup Analysis
- Sensitivity Analysis

Motivation

- Observational study to estimate the effect of a binary treatment T on a response Y, in the presence of a vector of covariates **X**.
- The propensity score e(**X**) is a device for constructing matched sets (of treatment and control subjects) or strata when **X** contains many covariates.

Statistical Definition

- Propensity score e(**x**) is the conditional probability of receiving the exposure given the observed covariates **x**.

e(**x**) =

$$\Pr(\text{Exposure} \mid X_{\text{subject}} = x)$$

Key Property of the Propensity Score

- Strata or matched sets that are homogeneous in e(**X**) tend to "balance" **X** -- treated and control subjects in the same stratum or matched set tend to have the same distribution of **X**.
- "Randomization" in an observational study (for observed covariates – unobserved then assumed unbalanced).

Estimating the Propensity Score: The Method

- Estimate a Logistic Regression Model:
  - Dependent Variable = Treatment Group (1 = treatment, 0 = control)
  - Independent Variables = observed covariates
  - Obtain Predicted Values for each subject – these are the estimated propensity scores

Choosing Variables for the Propensity Score Model

- PS should use all covariates that subject-matter experts (and subjects) judge important when selecting treatments.
- Also, use all covariates that relate to treatment and outcome, certainly including any covariate that improves prediction (of exposure group).
- Sop up as much “signal” as possible.

Assessing the Propensity Model

- Include all covariates of statistical or clinical relevance – non-parsimonious
- Main goal: We want to balance the covariates in the treatment groups
- We want to see good discrimination (high ROC area) in the final model
- Not concerned about over-fitting or external validity of the PS model.

Application 1: Matching on the Propensity Score

- Match subjects on many covariates with a single score
- Achieve covariate balance, given sufficient sample size.
- Forced check on whether covariate values overlap "enough"
- Emulate a RCT – "gold standard"

Matching: an example

- Problem: Is there a difference in 30 day mortality between ICU patients who received a Right Heart Catheter (RHC) and those who did not?
- Sample: 5735 ICU patients from 5 tertiary care hospitals; data collection 1989-1994
- Patients not randomized to treatments

Pre-Matched Data

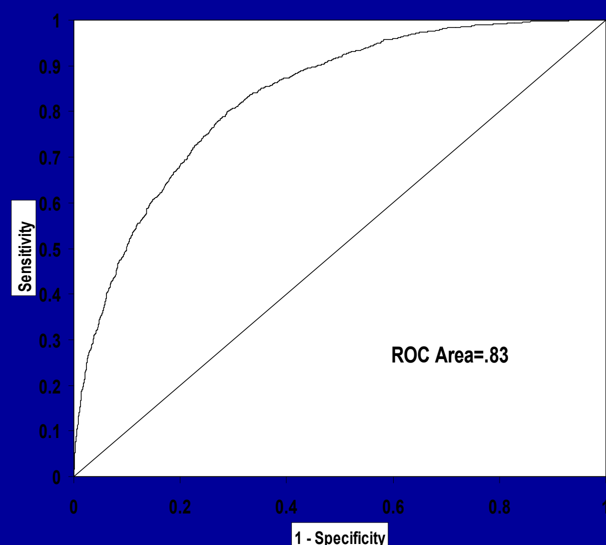
Variable		NO RHC	RHC	χ^2 or Z*	p
N		3551	2184		
Age, years	<50	25%	25%	55.7	.0001
	50 to <60	16%	17%	-	-
	60 to <70	23%	26%	-	-
	70 to <80	23%	24%	-	-
	>80	*14%*	*8%*	-	-
Disease Category	*ARF*	*34%*	*27%*	369.8	.0001
	MOSF	*35%*	*57%*	-	-
	CHF	*7%*	*10%*	-	-
	Other	24%	7%	-	-
DNR, day 1	*Yes*	*20%*	*14%*	46.2	.0001
APACHE III score (without coma score)		*51 (38-62)*	*61(47-74)*	18.1	.0001
Model estimate of the probability Of 2 month survival		0.61(0.49-0.76)	0.56(0.45-0.72)	-7.6	.0001
Physician estimate of the probability of 2 month survival		0.51 (0.10-0.80)	0.41(0.05-0.75)	-9.8	.0001
No. Comorbid Illnesses		*1.7(1-3)*	*1.6(1-3)*	-5.0	.0001
ADL 2 weeks prior		1.6(0.5-2-4)	1.5(0.5-2.2)	-1.6	.1039
DASI 2 weeks prior		20(16-23)	21(17-25)	3.2	.0013
LOS prior to study entry		4.6(0-5)	6.4(1-2)	-12.8	.0001

Estimated Probability of RHC

The Propensity Model to predict Y = use of a RHC

- "Main effects" logistic regression without interaction or higher order terms
- Predictors developed pre-data collection
- Predictors were: Age, Gender, Education, DNR, APACHE III, Number of Co-morbid Illnesses, Physiologic Measures
- Tried to maximize discrimination

ROC Area Shows Good Discrimination



"Nearest Available" Matching using Propensity Scores

- Randomly select a "treatment" patient.
- Match to the "control" patient with nearest propensity score.
- If no "control" patient has propensity score within .03 of "treatment" patient, no match.
- Repeat until no further matches are possible.
- Forced an equal number of matches with positive and negative propensity difference

Results of Matching

Variable	RHC	NoRHC	p
N	1008	1008	
Disease Category			1.000
ARF	*460 (45.6%)*	*460 (45.6%)*	
CHF	*113 (11.2%)*	*113 (11.2%)*	
MOSF	*340 (33.7%)*	*340 (33.7%)*	
Other	*95 (9.5%)*	*95 (9.5%)*	
Propensity for RHC	*0.51(0.35-0.67)*	*0.51(0.36-0.67)*	.8478
*APACHE III Score** (without coma score)	*57(44-71)*	*57(43-70)*	.3399
Model estimate of the probability of 2 month survival	0.58(0.46-0.74)	0.58(0.47-0.74)	.4294
Age, years	60(48-72)	60(49-73)	.9697
No. Comorbid Illnesses	*1.6(1-2)*	*1.6(1-2)*	.3999
ADL 2 weeks prior	1.5(0-2)	1.5(0-2)	.4258
DASI 2 weeks prior	21(16-24)	21(17-24)	.4802
LOS prior to study entry	6.8(0-8)	6.5(0-8)	.4559

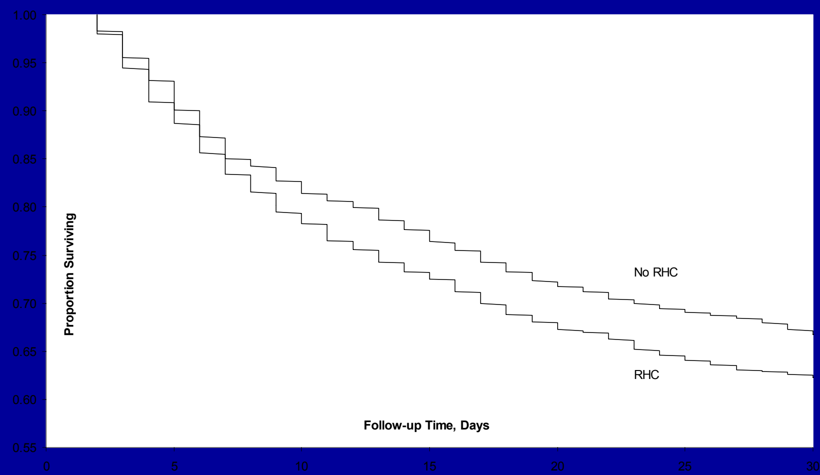
Results continued

- Results of matching: good covariate balance
- Conclusion: matching routine was successful

Matching: Testing Outcome

- McNemar's test of Binary outcome (30 day mortality) on treatment assignment showed χ^2 =4.49(p =.034)
- Kaplan-Meier log-rank test on survival to 30 days showed χ^2 = 4.72(p =.029)

Kaplan-Meier Curves of 30 Day Survival



Summary of Matching example

- Propensity model showed good discrimination
- 1008 matches obtained
- Symmetry was achieved
- Good covariate balance achieved
- Standard statistical tests used to test differences in outcome (30 day mortality)

Application 2: Subclassification on the Propensity Score

- Dividing observations into quintiles of propensity
- Subclassification on the propensity score balances all observed covariates
  - a generalization of subclassification with one covariate (e.g. age adjustment)
- Subclassification does not rely on a particular functional form (e.g. linearity)

Goals of Subclassification

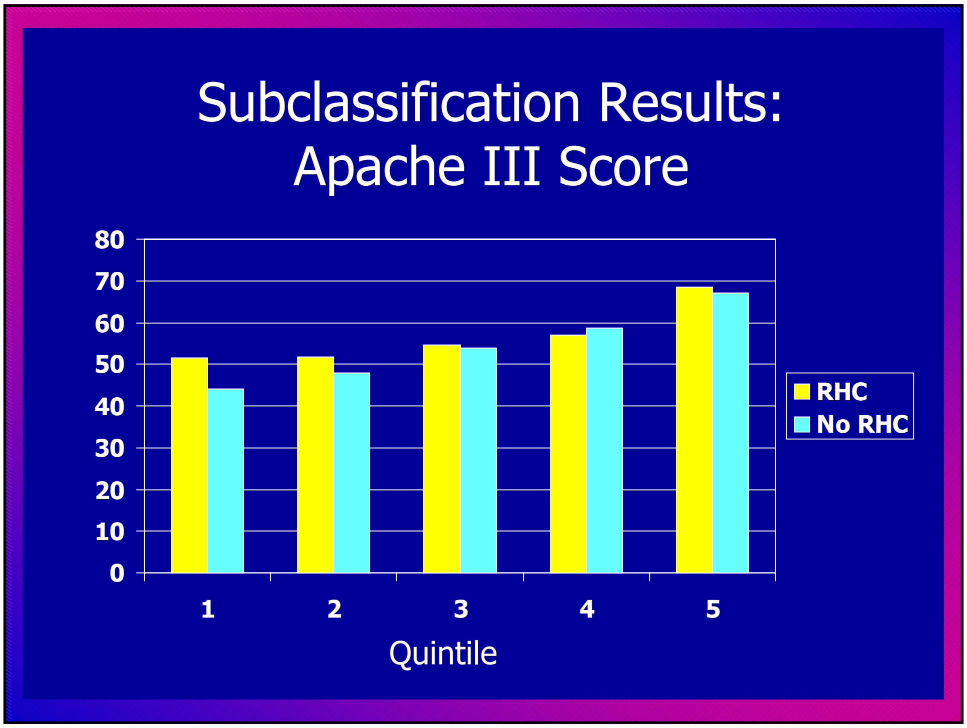
- Use entire Sample(versus matched subsample)
- Balance all observed covariates within subclass
- Directly compare RHC patients and non-RHC patients on 30 day survival

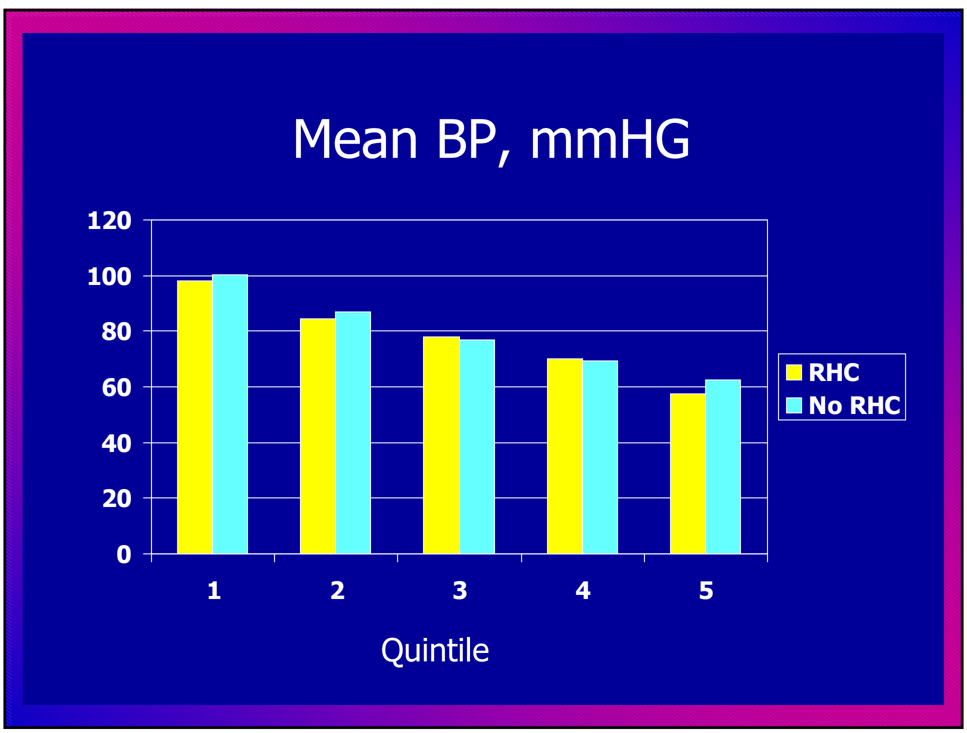
How to Subclassify

- Propensity score was estimated for 5,735 patients as previously shown
- Patients were then assigned to quintiles after patients were sorted on propensity

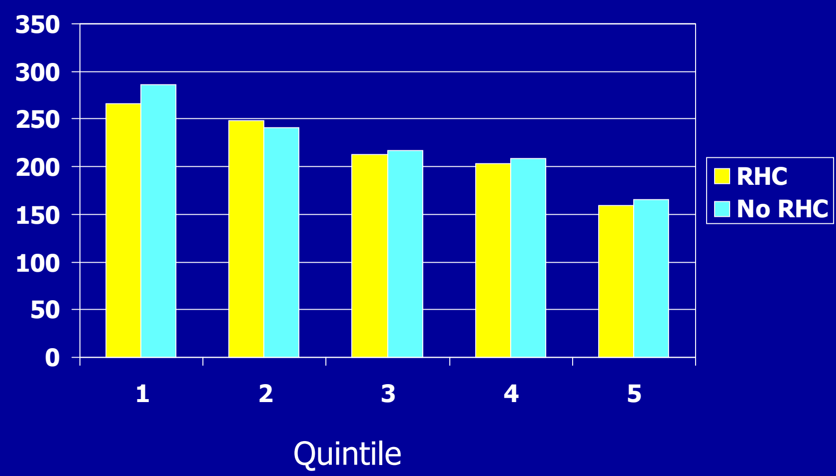
Sample size and range of propensity within quintile

Quintile	Range	RHC	No RHC
1	.001-.103	44	1103
2	.103-.239	203	944
3	.239-.439	404	743
4	.439-.663	627	520
5	.663-.988	906	241


Subclassification Results:
Apache III Score
80
70
60
50
40
30
20
10
0
1
2
3
4
5
Quintile
RHC
No RHC


Mean BP, mmHG
120
100
80
60
40
20
0
1
2
3
4
5
Quintile
RHC
No RHC

PaO2/FIO2, mmHG



Results of Subclassification

- 1100 observations within each subclass
- Good overlap between RHC and non-RHC patients in all subclasses

Subclassification: Testing Outcome

- Testing outcome can be done by estimating effect size for RHC patients and non-RHC patients within each subclass
- 30 day survival rates for RHC patients and non-RHC patients within each subclass can be assessed using Log Rank test or Likelihood Ratio test

Summary of Subclassification example

- Propensity model showed good discrimination
- 1100 patients in each subclass
- Good covariate balance achieved within each subclass
- Standard statistical tests could be used to directly compare RHC and non-RHC patients on 30 day survival

Application 3: Multivariate Regression with Propensity

- Evaluate association of treatment group with outcome using complete population
- Propensity score adjusts for all covariates in model
- Analyze subgroups (strata) using identical multivariate models

Regression Analysis using the Propensity Score

- Proportional Hazard model was estimated with survival to 30 days as the dependent variable
- Model was adjusted using the logit for selection to RHC management
- Clinical subgroups were separately analyzed using the identical model

Assessing Viability of the Propensity Score in Regression

- Divide population into subclasses of propensity and assess extent of covariate balance
- Standardized differences in covariate means
- Sensitivity analysis

Assessing Viability in our example

- We assessed balance by looking within subclasses (see subclassification example)
- Sensitivity analysis was conducted to address unobserved covariance

Multivariate Regression: Testing Outcome

- After adjustment for selection to RHC using the propensity score, RHC was associated with an increased risk of death(relative hazard of death=1.21;95 CI 1.09 - 1.25)
- Analysis of important clinical subgroups showed similar results

Summary of Multivariate Regression with Propensity Scores

- Propensity model showed good discrimination
- Logit of selection to RHC management was entered into a Proportional Hazards model
- Dependent variable was 30 day survival
- RHC was associated with an increased risk of death

Application 4: Analysis by Subgroup

- Estimate propensity score for each individual subgroup
- Alternative is to estimate propensity score for whole population and conduct analyses on individual subgroups
- Check for overlap and covariate balance within each subgroup

Analysis by Subgroup in RHC study

- Overlap and covariate balance checked within each important subgroup
- Subgroups examined were age group(elderly vs. young), gender, race, patients with shock or sepsis, patients receiving postoperative care, and disease group
- No subset where hazard ratio was significantly reduced by using RHC

Hidden Bias in Observational Studies

- Unmeasured confounders; uncorrelated with measured/adjusted confounders
- Sensitivity analysis
  - provides evidence about the degree to which study results are sensitive to hidden bias

Sensitivity Analysis

- Can be applied to most statistical methods:
  - Signed Rank, Rank Sum
  - Log-rank, McNemar
  - Cochran-Mantel-Haenzel
  - others
- "How much hidden bias would have to be present to alter the study's conclusions?"

Sensitivity Analysis: an example

- Estimate the effect of a possible missing covariate on adjustment using the propensity score
- How powerful would that unobserved covariate have to be to alter our conclusions

Sensitivity Analysis

Effects of an unobserved covariate on the probability of survival to 30 days for RHC patients and non-RHC patients

Effect of unobserved Covariate on group Membership	Effect of unobserved covariate on likelihood of survival for RHC Patients	Effect of unobserved covariate on likelihood of survival for non-RHC patients	Fraction of patients with unobserved covariate = 0
$\exp(\alpha)$	$\exp(\delta_0)$	$\exp(\delta_1)$	.5
Double the odds of being Managed with RHC	Halves the odds of survival	Halves the odds of survival	RHC No RHC
		Doubles the odds of survival	RHC No RHC
	Doubles the odds of survival	Halves the odds of survival	RHC No RHC
		Doubles the odds of survival	RHC No RHC

Conclusions of Sensitivity Analysis

- Missing covariate would have to increase the hazard* of death 6 fold
- AND
- Increase the probability of RHC 6 fold in order for a true relative relative hazard of .80 to be misrepresented as a relative hazard of 1.21

Summary: Sensitivity Analysis

- Question of hidden bias in observational studies
- Can be applied to many statistical tests
- In RHC study we assessed the ability of an unobserved covariate to change results
- We found that an unobserved covariate would have to be very powerful to alter conclusions

Summary

- Constructed a Propensity Score to counter effects of selection bias (selection to RHC)
- Three methods of applying the Propensity Score showed that good covariate balance was achieved
- Sensitivity analysis provided insight into the power of an unobserved covariate's ability to alter results