PubH 7405:
REGRESSION ANALYSIS

Propensity Score

INTRODUCTION:

There is a growing interest in using observational (or nonrandomized) studies to estimate the effects of treatments on outcomes. In observational studies, treatment selection is often influenced by subject characteristics. As a result, baseline characteristics of treated subjects often differ systematically from those of untreated subjects. Therefore, one must account for systematic differences in baseline characteristics between treated and untreated subjects when estimating the effect of treatment on outcomes.

Historically, applied researchers have relied on the use of regression adjustment to account for differences in measured baseline characteristics between treated and untreated subjects. Recently, there has been increasing interest in methods based on the propensity score to reduce or eliminate the effects of confounding when using observational data.

Since the seminal paper by Rosenbaum and Rubin (The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55; 1983) on propensity score analysis, research using propensity score analysis has grown exponentially over three decades. For example, a Google Scholar search for the phrase "propensity score" in articles published between 2000 and 2015 returned 63,200 results but only returned 868 results for articles published between 1985 and 1999.

Association of Perioperative β-Blockade With Mortality and Cardiovascular Morbidity Following Major Noncardiac Surgery

Martin J. London, MD
Kwan Hur, PhD
Gregory G. Schwartz, MD, PhD
William G. Henderson, PhD, MPH

JAMA 2013

Aim: To determine the association of early exposure to β-blockers with 30-day postoperative outcome in patients undergoing noncardiac surgery

Table 1. Selected Study Cohort Characteristics[a]

Variables	Full Cohort Percentage of Patients All (n = 136 745)	Exposed (n = 55 138)	Not Exposed (n = 81 607)	P Value
Age, mean (SD), y	64.4 (10.2)	65.3 (9.9)	61.3 (11.5)	<.001
Race				
White	66.4	67.5	65.1	<.001
Sex				
Men	96.3	96.9	94.1	<.001
Body mass index, mean (SD)[b]	29.3 (5.7)	29.3 (5.7)	28.8 (5.5)	<.001
Preoperative risk variables				
ASA physical status classification				
I	0.1	0.1	1.3	<.001
II	16.6	12.5	29.8	
III	74.4	75.7	63.5	
IV	8.7	11.6	5.3	
V	0.1	0.1	0.1	
Revised Cardiac Risk Index variables				
Congestive heart failure	9.8	15.2	5.3	<.001
Cerebrovascular disease	10.0	12.2	6.6	<.001
Diabetes (insulin)	10.4	12.5	6.5	<.001
Diabetes (insulin or oral)	27.5	30.7	19.1	<.001
Ischemic heart disease	28.0	42.9	14.7	<.001
High-risk surgery	40.4	40.9	35.9	<.001
Renal insufficiency	3.2	4.2	1.9	<.001
Laboratory measurements, mean (SD)				
Serum creatinine, mg/dL	1.2 (0.8)	1.2 (0.8)	1.1 (0.6)	<.001
Hematocrit, %	41.0 (8.8)	40.8 (4.9)	41.5 (4.6)	<.001
White blood cell count, $\times 10^3/\mu L$	7.9 (3.9)	8.0 (3.6)	7.8 (3.8)	<.001
Surgical details				
Surgery specialty				
General	26.6	26.1	24.3	<.001
Neurosurgery	8.6	7.9	10.2	
Orthopedic	28.7	25.9	33.5	
Otolaryngology	4.0	3.6	4.5	
Thoracic	6.9	7.2	5.1	
Urology	14.6	12.5	16.7	
Vascular	10.6	16.8	5.7	
Emergency procedure	6.4	6.2	6.6	.006
Laparoscopic procedure	10.8	10.7	10.1	<.001
Endovascular procedure	1.5	2.6	0.8	<.001
Duration of surgery, mean (SD), h	2.8 (1.7)	2.89 (1.7)	2.73 (1.7)	<.001
Principal anesthesia technique				
General	89.2	89.3	88.9	<.001
Spinal/epidural alone	8.9	8.6	9.3	
Other	1.9	2.1	1.8	

Different at baseline characteristics

You need to evaluate some treatment (Here: Exposed to β-blockade versus Non-exposed) but treatment assignment was influenced by patients' baseline characteristics (this is a non-randomized studies; treatment decided by doctors). Treatment groups are not similar; and many of these patients' baseline characteristics could be confounders or effect modifiers.

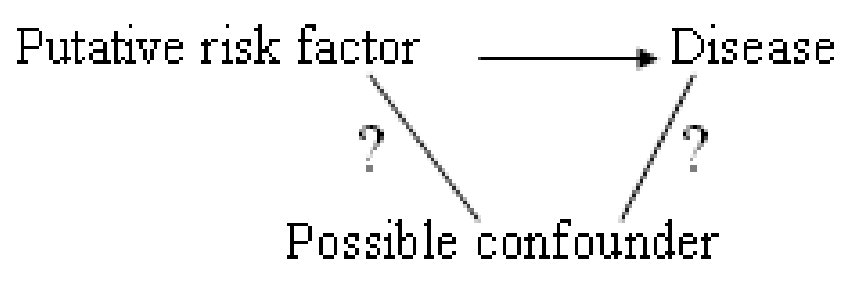
Confounding

- **Definition:** A situation in which the effect or association between an exposure (a predictor or risk factor) and outcome is distorted by the presence of another variable; this factor is not under investigation.
- A confounder meets all these conditions:

✓ It is a risk factor for the disease, independent of the putative risk factor.

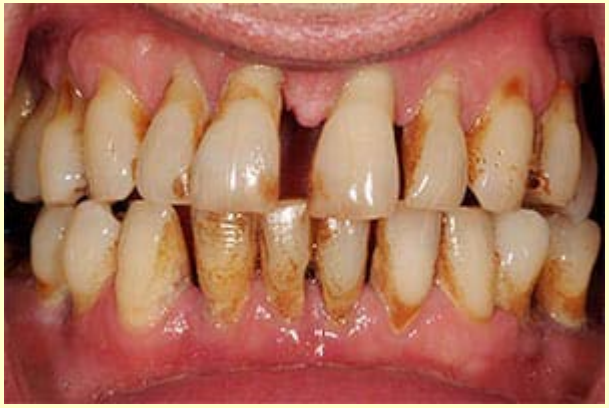
✓ It is associated with putative risk factor.

✓ It is not in the causal pathway between exposure and disease.



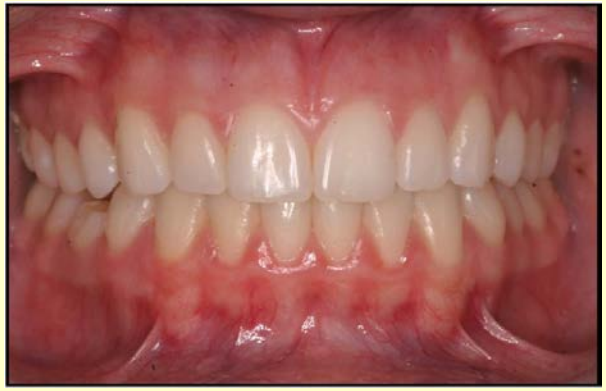
Example: Dental health vs. Heart disease

Periodontal disease



Cardiovascular disease

Periodontal health

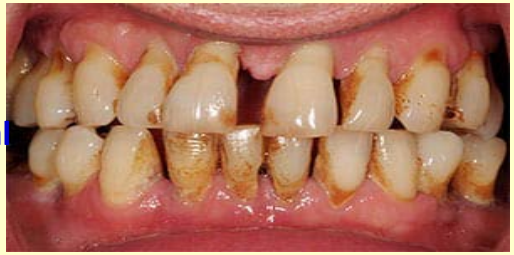


Cardiovascular disease

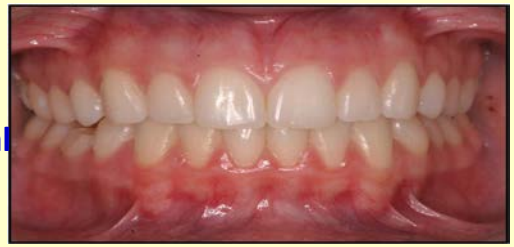
Example

We don't know whether the increase in the incidence is due to periodontal disease or smoking

Periodontal disease



Periodontal health



Cardiovascular disease

High incidence

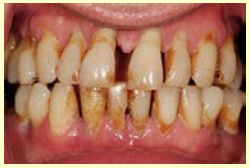
Cardiovascular disease

Low incidence

Smoking

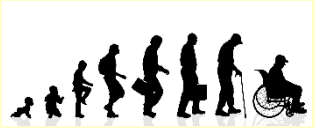
Confounding effects: Either way

- **Positive confounding**: overestimation of the true association



Smoking:
Positive confounder

- **Negative confounding**: underestimation of the true association



Age/Youth:
Negative confounder

Methods to address confounding

- Controlled in the **design phase**
  - Randomization
  - Restriction
  - Matching
  - Stratification

- Controlled in the **analysis phase**
  - Stratified analysis
  - Regression analysis

In Regression Analysis, assumptions are needed; some of these assumptions might not fit the data. It is more desirable to handle it in the Design Phase; but some ideal methods might not be possible.

RANDOMIZATION

- It removes bias in the treatment assignment
- It controls both known and unknown confounders
- It guarantees that statistical tests will have valid significance levels
- In short, it is the Gold Standard for clinical research designs.

RESTRICTION

- Exclusion of individuals with confounding factors or restriction to specific patient groups.
  - Example 1: Exclusion of smokers in the periodontal disease study
  - Example 2: Inclusion only males between 40-45 years
- Limitations:
  - ✓ Reduces the number of eligible individuals
  - ✓ Restriction limits generalizability
  - ✓ Inability to evaluate the effects of factors that been restricted for

This is similar to imposing Inclusion and Exclusion criteria in clinical trials but, in clinical trials, Inclusion and Exclusion Criteria are imposed before treatment assignment

MATCHING:

- Each pair of persons enrolled in a study are similar for one or more characteristics
  - Example: If a 60 year old Caucasian smoker with periodontal disease is entered then a 60 year old Caucasian smoker without periodontal disease will also be included
- Limitations:
  - ✓ Time-consuming and expensive
  - ✓ Limits sample size
  - ✓ Only for a limited number of confounding factors
  - ✓ Inability to evaluate the effect of the factors that have been matched

Very hard, even not possible, to match when there are many confounders or potential confounders; substantially reduced sample size.

STRATFICATION AND SUBGROUP ANALYSES:

- **Control for confounding by creating two or more categories or subgroups (strata) in which the confounding variable does not vary.**
  - **Example: Divide patients with and without periodontal disease into groups based on smoking status: smokers and non-smokers.**
- **Limitations:**
  - ✓ **Inability to control simultaneously for multiple confounding variables**
  - ✓ **Limits sample size**
  - ✓ **Time-consuming**

In practice, though Randomized Clinical Trial is the best choice, sometimes it is hard to carry out. Why? A variety of reasons:

- **Unethical**
- **Infeasible**
- **Impractical**
- **Not scientifically or financially justified**

Observational studies are always easy to carry out, but hard to draw causal inference.

Treatment selection is influenced by subject characteristics, which is a nonrandomized and uncontrolled process.

How to account for this systematic difference?

There is a method which has been grown in popularity. It creates, post hoc, similar groups – somewhat similar to post hoc stratification. It is an application of Logistic Regression, and called “Propensity Score”.

Propensity score:

The probability of a unit being assigned to treatment group conditional on observed baseline covariates.

Rosebaum and Rubin (1983)

IMPLEMENTATION:

(Multiple) Logistic Regression with
(1) Dependent Variable: Treatment assignment: Z = 1 for a subject in treatment group; Z = 0 for a subject in control group (We reserve the notation Y for outcome);
(2) Covariates (X's) include all possible confounders – including baseline characteristics, among others.

Choosing variables and performing calculation

- Choose relevant covariates
- Run a logistic regression
  - Treatment group coded 1, control coded 0
  - Prob(Z=1 | X_1, X_2, X_3, … X_n) = Propensity Score

RATIONALE FOR USE:

If units have the same propensity score (probability to be assigned to the treatment group, Pr(Z=1), determined by the logistic regression model), they will have the same covariate values.

The propensity score is the probability of treatment assignment conditional on observed baseline characteristics. The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics of a randomized controlled trial. In particular, the propensity score is a balancing score: conditional on the propensity score, the distribution of observed baseline covariates will be similar between treated and untreated subjects.

Methods for Propensity Score:

There are 3 simple methods using propensity score: matching on the propensity score, stratification on the propensity score, and covariate adjustment using the propensity score:

- **Matching**
- **Stratification**
- **Covariate Adjustment**

Propensity Score Matching

Propensity score matching entails forming matched sets of treated and untreated subjects who share a similar value (up to some level of difference) of the propensity score. The most common implementation of propensity score matching is one-to-one or pair matching. Once a matched sample has been formed, the treatment effect can be estimated by directly comparing outcomes between treated and untreated subjects in the matched pair.

(1) If the outcome is continuous (e.g., a depression scale), the effect of treatment can be estimated as the difference between the mean outcome for treated subjects and the mean outcome for untreated subjects; e.g. One-sample t-test.

(2) If the outcome is binary, the effect of treatment can be estimated as the difference between the proportion of subjects experiencing the event in each of the two groups (treated vs. untreated); e.g. McNemar Chi-square test.

Matched pairs are often formed without replacement; and there are two different ways to achieve balanced matched samples: greedy or optimal matching. The processes are different but the results are similar.

In greedy matching, a treated subject is first selected at random. The untreated subject whose propensity score is closest to that of this randomly selected treated subject is chosen for matching to this treated subject. This process is then repeated until one has exhausted the list of treated subjects. This process is called greedy because at each step in the process, the nearest untreated subject is selected for matching to the given treated subject, even if that untreated subject would better serve as a match for a subsequent treated subject.

An alternative to greedy matching is optimal matching, in which matches are formed so as to minimize the total within-pair difference of the propensity score. Gu and Rosenbaum (Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics 2, 405–420; 1993) compared greedy and optimal matching and found that optimal matching did no better than greedy matching in producing balanced matched samples.

Stratification on the Propensity Score

Stratification on the propensity score involves stratifying subjects into mutually exclusive subsets based on their estimated propensity scores.

Subjects are ranked according to their estimated propensity scores; then stratified into subsets based on previously defined thresholds of the estimated propensity score.

A common approach is to divide subjects into five equal-size groups using the quintiles of the estimated propensity score.

Cochran (The effectiveness of adjustment by sub-classification in removing bias in observational studies; Biometrics 24, 295–313; 1968) demonstrated that stratifying on the quintiles of a continuous confounding variable eliminated approximately 90% of the bias due to that variable. Rosenbaum and Rubin (Reducing bias in observational studies using sub-classification on the propensity score; Journal of the American Statistical Association 79, 516–524; 1984) extended this result to stratification on the propensity score, stating that stratifying on the quintiles of the propensity score eliminates approximately 90% of the bias due to measured confounders when estimating a linear treatment effect.

Stratification on the propensity can be conceptualized as a meta-analysis of a set of (five) randomized clinical trials. Within each stratum, the effect of treatment on outcomes can be estimated by comparing outcomes directly between treated and untreated subjects. The stratum-specific estimates of treatment effect can then be pooled across stratum to estimate an overall treatment effect; e.g. using weighted average.

Another propensity score method is covariate adjustment using the propensity score. Using this approach, the outcome variable is regressed on an indicator variable denoting treatment status and the estimated propensity score. The choice of regression model would depend on the nature of the outcome. For continuous outcomes, a normal error linear model would be chosen; for dichotomous outcomes, a logistic regression model may be selected. The effect of treatment is determined using the estimated regression coefficient from the fitted regression model.

For a linear model, the treatment effect is an adjusted difference in means, whereas for a logistic model it is an adjusted odds ratio. Of the three simple propensity score methods, this is the only one that requires that a regression model relating the outcome to treatment status and a covariate (the propensity score) be specified. Furthermore, this method assumes that the nature of the relationship between the propensity score and the outcome has been correctly modeled.

This is similar to using regression, without propensity score, to adjust for differences in baseline characteristics. But instead of using multiple predictors, all baseline characteristics are combined (through propensity score) into one “index” (the propensity score). This makes it more simple to check for model assumptions.

Comparison of the Different Propensity Score Methods:

Several studies have demonstrated that propensity score matching eliminates a greater proportion of the systematic differences in baseline characteristics between treated and untreated subjects than does stratification on the propensity score or covariate adjustment using the propensity score (e.g. Austin, P. C. Type I error rates, coverage of confidence intervals, and variance estimation in propensity-score matched analyses. The International Journal of Biostatistics, 5, Article 13; 2009). However, it's more time-consuming and limits sample size (some subjects were not able to match)

BALANCE DIAGNOSTICS

The true propensity score is a balancing score: conditional on the true propensity score, the distribution of measured baseline covariates is independent of treatment assignment. In an observational study the true propensity score is not known. It must be estimated using the study data. An important component of any propensity score analysis is examining whether the propensity score model has been adequately specified.

With propensity score matching, assessing whether the propensity score model has been adequately specified involves comparing treated and untreated subjects within the propensity score matched pairs. For stratification on the propensity score, this assessment entails comparing treated and untreated subjects within strata of the propensity score.

VARIABLE SELECTION FOR THE PROPENSITY SCORE MODEL

There is a lack of consensus in the applied literature as to which variables to include in the propensity score model. Possible sets of variables for inclusion in the propensity score model include the following: (1) all measured baseline covariates, (2) all baseline covariates that are associated with treatment assignment, (3) all covariates that affect the outcome (i.e., the potential confounders), and (4) all covariates that affect both treatment assignment and the outcome (i.e., the true confounders).

The propensity score is defined to be the probability of treatment assignment. Thus, there are theoretical arguments in favor of the inclusion of only those variables that affect treatment assignment (#2).

A recent study (Austin, Grootendorst, & Anderson. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study; Statistics in Medicine 26, 734–753; 2007) examined the relative benefits of including the different sets of baseline covariates described earlier in the propensity score model. It was shown that there were merits to including only the potential confounders or the true confounders in the propensity score model (#3 and #4).

It should be noted that, in practice, it may be difficult to accurately classify baseline variables into the true confounders, those that only affect the outcome, those that only affect exposure, and those that affect neither treatment nor the outcome. In specific settings, the published literature may provide some guidance for identifying variables that affect the outcome. In practice, in many settings, most subject-level baseline covariates likely affect both treatment assignment and the outcome. Therefore, in many settings, it is likely that one can safely include all measured baseline characteristics in the propensity score model.

If balance diagnostics are successful, does this make Propensity Score as good as Randomization?

A sort answer is “NO”.

With Propensity Score, one can only account for observed covariates in propensity score models whereas Randomization controls both known and unknown confounders

Regression Adjustment Versus Propensity Score:

There are good reasons for preferring the use of propensity score-based methods to regression-based methods when estimating treatment effects using observational data. It is simpler to determine whether the propensity score model has been adequately specified than to assess whether the regression model relating treatment assignment and baseline covariates to the outcome has been correctly specified.

However, with propensity score, one can only account for observed covariates used in propensity score models; there is a concern that it might make some unobserved confounder or confounders even more unbalanced. There are methods to check for this phenomenon, collectively called sensitivity analysis that assesses the potential impact of unobserved confounders on the treatment effect but they are all rather complicated (e.g. Li, Chen et. al. Propensity Score-based Sensitive Analysis Method for Uncontrolled Confounding. American Journal of Epidemiology 174: 345-353; 2011).

TREATMENT EFFECT

After Propensity Scores are generated, it is possible to estimate the treatment effect;

(1) If Matching is used, one can estimate within each pair, the average them out;

(2) If Stratification is used, one can similarly estimate within each stratum, then weighted average them out;

(3) If Covariate Adjustment is used, the effect of treatment is determined using the estimated regression coefficient from the fitted regression model.

There are other methods, collectively called Inverse Probability of Treatment Weighting (IPTW). Lunceford and Davidian (2004; Stratification and Weighting via propensity Score in estimation of causal treatment effects: A comparative study; Statistics in medicine 9: 403-425) review a variety of estimators for treatment effects based on IPTW. Assume that Y denotes the outcome variable, an estimate of the ATE is:

$$ATE = \frac{1}{n}\sum\frac{ZY}{e} - \frac{1}{n}\sum\frac{(1-Z)Y}{1-e}$$

where Z is the treatment, e the propensity score, and n denotes the number of subjects

EXAMPLE

CONNORS JR AF, SPEROFF T, DAWSON NV, ET AL. (1996) "THE EFFECTIVENESS OF RIGHT HEART CATHETERIZATION (RHC) IN THE INITIAL CARE OF CRITICALLY ILL PATIENTS."
***JAMA*, 276(11): 889-897**

SPECIFIC AIM:

- A prospective cohort study that examined the association between right heart catheterization (RHC) during the first 24 hours in the ICU and survival time, length of hospital stay, intensity of care, and cost of care.
- Use of propensity scores and case-matching to adjust for treatment selection bias

BACKGROUND:

- Many physicians consider RHC as a direct measurement of cardiac function necessary to guide therapy decision-making and management for critically ill patients. These physicians believe that the use of RHC leads to better outcomes for the patients.
- The benefit of RHC (as of publication date 1996) has not been demonstrated in a randomized controlled trial due to ethical considerations. Physicians refuse to allow patients to be randomized out of concern for those assigned to the control group.

- Only option is observational study without randomization of patients into case and control groups.
- Physicians make treatment decisions based on patient factors that are also related to outcomes of interest.
- In observational studies, the decision-making process that results in the creation of the RHC and non-RHC patient groups creates treatment selection bias.

Data:

- Five teaching hospitals in the US between 1989 and 1994
- Total participants: 5735 critically ill adult patients receiving care in ICU for 1 of 9 disease categories - acute respiratory failure (ARF), chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), cirrhosis, nontraumatic coma, colon cancer metastatic to the liver, non-small cell cancer of the lung, and multiorgan system failure (MOSF) with malignancy or sepsis.
- Patients were followed up for 6 months.

DATA COLLECTION & PREPARATION

- Determined whether RHC was used within first 24 hours according to chart documentation. Total RHC patients = 2184. Non-RHC patients = 3551.
- Chart and other hospital documentation analyzed to identify patients meeting entry criteria, disease diagnosis, physiological status, intensity of care, length of stay, and total cost of care.
- Interviews conducted to gather demographic information.

Variables

- **Outcome variables:**
  - Survival at 30 days, 2 months, and 6 months
  - Hospital length of stay
  - Cost of care
  - Intensity of care
- **Predictor variable**
  - RHC/no RHC
- **Covariates**
  - Demographic including race, sex, age, disease category, cancer status, insurance status
  - Variety of physiological variables

Analysis Methods

- Propensity scores for RHC constructed using multivariable logistic regression.
- Case-matching Method used to estimate association of RHC with outcomes of interest after adjusting for treatment selection using propensity score.

Creating Propensity Scores

- Physician-identified variables that relate to the decision to use or not to use RHC.
- Logistic regression analysis performed as follows:
  - Dependent Variable: RHC or No RHC
  - Independent Variables: age, sex, race, education, income, type of insurance, disease category, admission diagnosis, ADL and DASI, DNR status, cancer status, physiology components
- From this model they obtained Pr(RHC) = propensity score for each patient (probability to receive RHC).
- Stratifying by quintiles of propensity for RHC.

Case-matching

- RHC matched to no-RHC based on propensity score and disease category
- Randomly select RHC patient, then match to no-RHC with same disease category who had most similar propensity score (within 0.03)
- Continued until all pairs identified

Results

Unadjusted: Patients with RHC had an increased mortality, higher mean hospital costs, longer length of stay

Outcome	No RHC (n=3551)	RHC (n=2184)	P
Survival, No. (%)			
30 d	2463 (69.4)	1354 (62.0)	<.001
2 mo	2231 (62.8)	1190 (54.5)	<.001
6 mo	1906 (53.7)	1012 (46.3)	.001
Resource utilization†			
Total costs (× $1000)‡	74.3 [18.4, 37.1, 81.5]	131.9 [42.1, 81.7, 160.6]	<.001
Average TISS	28 [21, 27, 35]	35 [28, 35, 42]	<.001
Length of stay, d†			
ICU	10.3 [3, 6, 11]	15.5 [5, 9, 18]	<.001
Study	20.5 [8, 13, 23]	25.7 [9, 17, 32]	<.001

*TISS indicates Therapeutic Intervention Scoring System (with points associated with RHC removed); and ICU, intensive care unit.

†Continuous variables are presented as mean [25th, 50th (median), and 75th percentiles].

‡Total costs are estimated hospital costs from study day 1 to discharge (see "Methods" for details).

Adjustment for Treatment Selection Bias

- 1008 successfully matched pairs
- No differences within these pairs for 18 variables pertaining to health status.

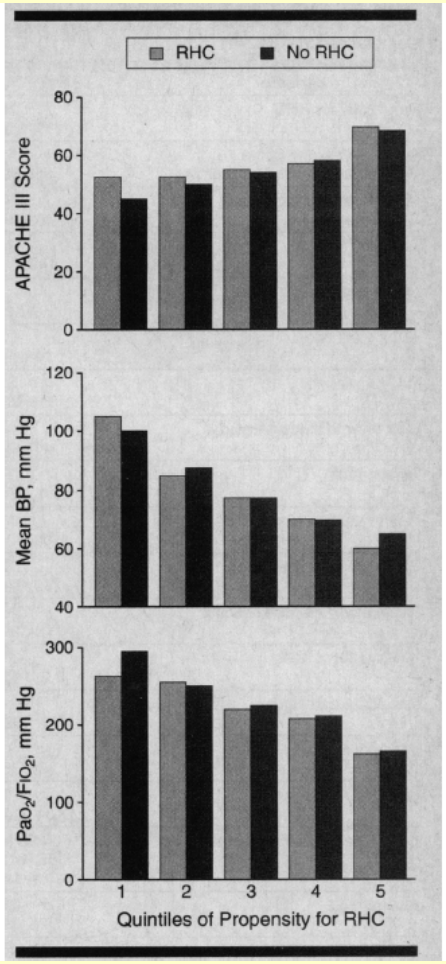
Table 3.—Characteristics of 1008 Matched Pairs of Patients Managed With and Without Right Heart Catheterization (RHC)*

Variable	No RHC (n=1008)	RHC Day 1 (n=1008)	P
Propensity for RHC	0.51 [0.35, 0.50, 0.67]	0.51 [0.36, 0.50, 0.67]	.85
Acute Physiology Score (without Glasgow Coma Score)	57 [44, 58, 71]	57 [43, 57, 70]	.34
Model estimate, probability of 2-mo survival	0.58 [0.46, 0.62, 0.74]	0.59 [0.47, 0.62, 0.74]	.43
Age, y	60 [49, 63, 72]	60 [49, 62, 73]	.97
No. of comorbid illnesses	1.6 [1, 1, 2]	1.6 [1, 1, 2]	.40
ADLs 2 wk prior	1.5 [0, 1, 2]	1.5 [0, 2, 2]	.43
DASI 2 wk prior	21 [16, 20, 24]	21 [17, 20, 24]	.48
LOS prior to study entry, d	6.8 [0, 2, 8]	6.5 [0, 2, 8]	.46
Temperature, °C	37.7 [36.1, 38.3, 39.1]	37.7 [36.2, 38.2, 39.0]	.92
Heart rate, beats/min	111 [105, 125, 145]	111 [103, 124, 145]	.75
Blood pressure, mm HG	73 [49, 61, 108]	71 [49, 60, 81]	.60
Respiratory rate, breaths/min	28 [19, 30, 38]	28 [14, 28, 38]	.30
WBC count × 10^9L	15.3 [8.2, 14.0, 20.0]	15.0 [7.4, 13.6, 20.0]	.53
PaO_2/FIO_2, mm Hg	210 [127, 185.7, 296]	211 [120, 192, 305]	.70
$PaCO_2$, mm Hg	37 [31, 36, 41]	38 [31, 36, 40]	.80
pH	7.39 [7.34, 7.40, 7.46]	7.39 [7.34, 7.40, 7.46]	.74
Creatinine, μmol/L (mg/dL)	203 (2.3) [88, 141, 230] (0.1, 1.6, 2.6)	203 (2.3) [106, 150, 239] (1.2, 1.7, 2.7)	.38
Albumin, g/L	30 [25, 35, 35]	30 [26, 35, 35]	.77
Glasgow Coma Score	13 [12, 15, 15]	13 [12, 15, 15]	.35

*Continuous variables are presented as mean [25th, 50th (median), 75th percentiles]. See Table 1 for explanation of abbreviations.

Multivariable regression for propensity for RHC

- Good discrimination between RHC and non-RHC patients:
- area under the receiver operating characteristic curve of 0.83.
- RHC mean propensity score: 0.577
- non-RHC mean propensity score: 0.253.
- Within pair differences in covariates within quintiles of propensity for RHC: not significantly different.



Adjusted: RHC survival lower at 30, 60, and 180 days after study entry.

Table 4.—Relationship of Right Heart Catheterization (RHC) to Survival for Matched Pairs of Patients Managed With and Without RHC*

Survival Interval	Survival, No. (%)		OR (95% CI)	P
	No RHC (n=1008)	RHC (n=1008)		
30 d	677 (67.2)	630 (62.5)	1.24 (1.03-1.49)	.03
60 d	604 (59.9)	550 (54.6)	1.26 (1.05-1.52)	.01
180 d	522 (51.2)	464 (46.0)	1.27 (1.06-1.52)	.009
Hospital	629 (63.4)	565 (56.1)	1.39 (1.15-1.67)	.001

*OR indicates odds ratio; and CI, confidence interval.

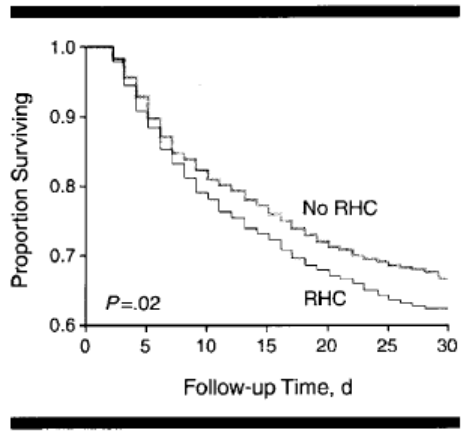


Figure 2.—Thirty-day survival curves for 2016 patients with and without right heart catheterization (RHC) matched for disease category and propensity score (case-matched analysis). Proportion of patients surviving with RHC and without RHC are shown over the 30 days after study entry. Survival is significantly better in the population managed without RHC.

Adjusted: RHC patients associated with higher costs of care, higher intensity of care, and longer stay in ICU compared to non-RHC patients

Table 5.—Relationship of Right Heart Catheterization (RHC) to Resource Use for Matched Pairs of Patients Managed With and Without RHC*

	No RHC (n=1008)	RHC (n=1008)	P
Resource utilization			
Total costs (× $1000)†	35.7 (11.3, 20.6, 39.2)	49.3 (17.0, 30.5, 56.6)	<.001
Average TISS (adjusted)	30 (23, 29, 38)	34 (27, 34, 41)	<.001
Length of stay, d			
ICU	13.0 (4, 7, 14)	14.8 (5, 9, 17)	<.001
Study	23.8 (9, 15, 28)	25.1 (9, 16, 31)	.14

*Continuous variables are presented as mean (25th, 50th [median], 75th percentiles). TISS indicates Therapeutic Intervention Scoring System (with points associated with RHC removed); and ICU, intensive care unit.
†Total costs are estimated hospital costs from study day 1 to discharge (see "Methods" for details).

Multivariable Analysis Results

- RHC vs. no-RHC: Increase risk of death, increase cost, increase length of stay
- Within clinical subgroups, there is no evidence that RHC is associated with decrease in relative hazard of death or increase in patient outcomes.

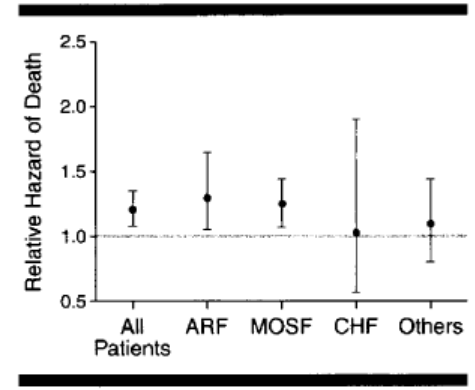


Figure 3.—Association of right heart catheterization with survival time, overall and within disease categories. The relative hazard of death associated with right

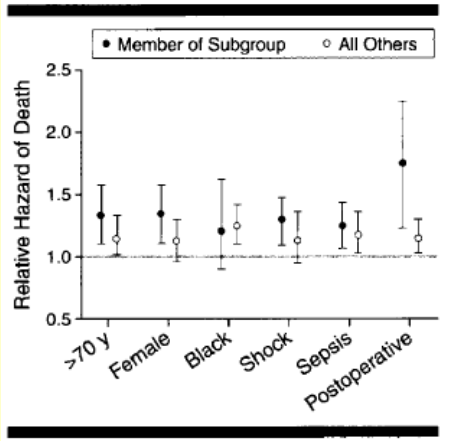


Figure 4.—Association of right heart catheterization with survival time in important subgroups. The ad-

Conclusions:

- After adjustment for treatment selection bias – through the use of propensity Scores, RHC is associated with a decrease in survival and an increase in cost and intensity of care.
- Suggests that RHC use should be reexamined through additional observational studies to assess the procedure's usefulness in terms of patient outcomes.

We can even apply the Propensity Score Method to small and mid-size Clinical Trials.

- We commonly check to see if randomization works by checking each variable, baseline measures and demographic characteristics, was balanced out.
- Factor or factors found not quite balanced would be included in the analysis as covariates in regression model.

- Instead of checking each variable individually, we could generate propensity scores (probabilities to get assigned to one of the two groups), then comparing the distributions of propensity scored (means or histograms).
- If randomization did not work very well, propensity score, or some function of propensity score, would be included as a covariate in the regression model

RECOMMENDED READING

The central role of the propensity score in observational studies for causal effects.
***Biometrika 70*, 41–55; 1983**

Due As Homework

#25.1 We have a data set on 191 Head and Neck cancer patients who were treated by radiation (File: Radiation). Data include Age, Gender, Income, Tumor site (1-6; e.g.1 = oral cavity, 3 = salivary gland), Stage T, N, M (0-4), a number of co-morbidities (diabetes, heart, stroke, lunch, arthritis; 0 = no, 1 = yes), smoking, drinking (0 = never, 1 = former, 2 = current), psychological illness (0 = no, 1 = yes), chemotherapy (ctx; 0 = no, 1 = yes). The last item, ACADEME, indicates if the patient treated at an academic health center (0 = no, 1 = yes – either U of M, Mayo Clinic).

a) Use the last item as a dependent variable in an application of Logistic Regression to investigate if it is significantly related to any factor or factors (Note: the common claim is that Academic Health Centers usually admit and treat more severe cases – if true, comparisons of outcomes should be adjusted accordingly).

b) Choose the factor which are significant at the 10% level (p-value < .10), refit the logistic model, calculate the propensity score for each patient; use these scores to compare the mean of two groups (academic and non-academic) using the two-sample t-test.

#25.2 We have data for 233 throat cancer patients (File; Throat-Cancer). Data include disease stages (T-stage and N-stage), co-morbidity score, smoking (packs/day), drinking (drink/day), and tumor site (should be regrouped into 3 categories: 1480-1489, 1610, and 1611-1619). This is an observational study, treatment is assigned by the patient's physician, not randomized. The treatment assignment might depend on patient's characteristics, disease condition, and/or preference. Some was treated by surgery (Treatment = 1), some by radiation and/or chemotherapy (treatment = 2). Two outcomes are: Recurrence and Q-of-L score.

(1) Use logistic regression, with Treatment as the dependent variable, to calculate Propensity score for each subject in both groups;

(2) Compare the average Q-of-L score of the two treatment groups, adjusted for propensity score found in question (1).