Propensity Scores in Medical Research

Robert H. Habib

November 11, 2013
Objectives

✓ Briefly review time-line of using propensity scores in medical research

✓ Learn what is a propensity score and How it can be used to analyze and interpret comparative data (Treatments; Risk Groups)

✓ Focus on use of Propensity Matching to simulate Randomized comparisons

✓ Discuss Propensity Matching Examples in Cardiovascular Medicine
The central role of the propensity score in observational studies for causal effects

By PAUL R. ROSENBAUM

Departments of Statistics and Human Oncology, University of Wisconsin, Madison, Wisconsin, U.S.A.

AND DONALD B. RUBIN

University of Chicago, Chicago, Illinois, U.S.A.

SUMMARY

The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates. Both large and small sample theory show that adjustment for the scalar propensity score is sufficient to remove bias due to all observed covariates. Applications include: (i) matched sampling on the univariate propensity score, which is a generalization of discriminant matching, (ii) multivariate adjustment by subclassification on the propensity score where the same subclasses are used to estimate treatment effects for all outcome variables and in all subpopulations, and (iii) visual representation of multivariate covariance adjustment by a two-dimensional plot.
Timeline: Use of Propensity Scores

Search Term:
(propensity score OR (propensit* AND (score* OR scoring* OR match* OR adjust*))))
The rate of using propensity scores in Pubmed articles has increased at about **130 times** the rate of increased in articles overall.
Why the Fast Rising Popularity of Propensity Scores?
Levels of Evidence

Entirely Dependent on below

Great, But expensive, time consuming, often impractical (or not doable)
Attractive features of RCTs

Randomization in RCTs

Treatment and Control study arms are similar in terms of observed and unobserved characteristics.

Any average difference in outcomes between the groups can only be attributed to the Treatment effect.

And, hence, quantifying the Treatment effect is simple, rarely requires risk adjustment or regression modeling.
Bardoxolone Methyl in Type 2 Diabetes and Stage 4 Chronic Kidney Disease

Dick de Zeeuw, M.D., Ph.D., Tadao Akizawa, M.D., Ph.D.,
Paul Audhya, M.D., M.B.A., George L. Bakris, M.D., Melanie Chin, Ph.D.,
Heidi Christ-Schmidt, M.S.E., Angie Goldsberry, M.S., Mark Houser, M.D.,
Melissa Krauth, M.B.A., Hiddo J. Lambers Heerspink, Pharm.D., Ph.D.,
Giuseppe Remuzzi, M.D., Robert D. Toto, M.D., Nosratola D. Vaziri, M.D.,
Christoph Wanner, M.D., Janet Wittes, Ph.D., Danielle Wrolstad, M.S.,
and Glenn M. Chertow, M.D., M.P.H., for the BEACON Trial Investigators*

This article was published on November 9, 2013, at NEJM.org.

DOI: 10.1056/NEJMoa1306033
Copyright © 2013 Massachusetts Medical Society.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Placebo (N = 1097)</th>
<th>Bardoxolone Methyl (N = 1088)</th>
<th>All Patients (N = 2185)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age — yr</td>
<td>68.2±9.4</td>
<td>68.9±9.7</td>
<td>68.5±9.6</td>
</tr>
<tr>
<td>Female sex — no. (%)</td>
<td>472 (43)</td>
<td>462 (42)</td>
<td>934 (43)</td>
</tr>
<tr>
<td>Race — no. (%)†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>848 (77)</td>
<td>846 (78)</td>
<td>1694 (78)</td>
</tr>
<tr>
<td>Black</td>
<td>176 (16)</td>
<td>185 (17)</td>
<td>361 (17)</td>
</tr>
<tr>
<td>Other</td>
<td>73 (7)</td>
<td>57 (5)</td>
<td>130 (6)</td>
</tr>
<tr>
<td>Hispanic ethnic group — no. (%)†</td>
<td>184 (17)</td>
<td>186 (17)</td>
<td>370 (17)</td>
</tr>
<tr>
<td>Region — no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>772 (70)</td>
<td>773 (71)</td>
<td>1545 (71)</td>
</tr>
<tr>
<td>European Union</td>
<td>169 (15)</td>
<td>165 (15)</td>
<td>334 (15)</td>
</tr>
<tr>
<td>Australia</td>
<td>67 (6)</td>
<td>66 (6)</td>
<td>133 (6)</td>
</tr>
<tr>
<td>Canada</td>
<td>46 (4)</td>
<td>41 (4)</td>
<td>87 (4)</td>
</tr>
<tr>
<td>Israel</td>
<td>23 (2)</td>
<td>23 (2)</td>
<td>46 (2)</td>
</tr>
<tr>
<td>Mexico</td>
<td>20 (2)</td>
<td>20 (2)</td>
<td>40 (2)</td>
</tr>
<tr>
<td>Weight — kg</td>
<td>95.3±21.1</td>
<td>95.1±22.0</td>
<td>95.2±21.5</td>
</tr>
<tr>
<td>Glycated hemoglobin — %</td>
<td>7.1±1.2</td>
<td>7.2±1.3</td>
<td>7.1±1.2</td>
</tr>
<tr>
<td>Serum creatinine — mg/dl</td>
<td>2.7±0.6</td>
<td>2.7±0.6</td>
<td>2.7±0.6</td>
</tr>
<tr>
<td>Estimated GFR — ml/min/1.73 m²</td>
<td>22.5±4.6</td>
<td>22.4±4.3</td>
<td>22.5±4.5</td>
</tr>
<tr>
<td>Urinary albumin-to-creatinine ratio‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>351</td>
<td>292</td>
<td>320</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>60–1136</td>
<td>53–1151</td>
<td>57–1140</td>
</tr>
<tr>
<td>Diabetes history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time since diabetes diagnosis — yr</td>
<td>18.1±9.7</td>
<td>18.9±9.7</td>
<td>18.5±9.7</td>
</tr>
<tr>
<td>Retinopathy — no. (%)</td>
<td>445 (41)</td>
<td>446 (41)</td>
<td>891 (41)</td>
</tr>
<tr>
<td>Neuropathy — no. (%)</td>
<td>500 (46)</td>
<td>517 (48)</td>
<td>1017 (47)</td>
</tr>
<tr>
<td>Amputation — no. (%)</td>
<td>49 (4)</td>
<td>57 (5)</td>
<td>106 (5)</td>
</tr>
<tr>
<td>Foot ulcer — no./total no. (%)</td>
<td>37/428 (9)</td>
<td>36/433 (8)</td>
<td>73/861 (8)</td>
</tr>
<tr>
<td>Cardiovascular history — no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any history of cardiovascular disease</td>
<td>619 (56)</td>
<td>609 (56)</td>
<td>1228 (56)</td>
</tr>
<tr>
<td>Hospitalization for heart failure</td>
<td>118 (11)</td>
<td>120 (11)</td>
<td>238 (11)</td>
</tr>
</tbody>
</table>
Levels of Evidence

- Editorials, Expert Opinion
- Case Series, Case Reports
- Case-Control Studies
- Cohort Studies
- Randomized Controlled Trials
- Systematic Reviews

Entirely Dependent on below

Great, But expensive, time consuming, often impractical (or not doable)

Observational, convenient but little control ...
TUTORIAL IN BIOSTATISTICS

PROPENSITY SCORE METHODS FOR BIAS REDUCTION IN THE COMPARISON OF A TREATMENT TO A NON-RANDOMIZED CONTROL GROUP

RALPH B. D'AGOSTINO, Jr.*

Department of Public Health Sciences, Section on Biostatistics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157-1063, U.S.A.

SUMMARY

In observational studies, investigators have no control over the treatment assignment. The treated and non-treated (that is, control) groups may have large differences on their observed covariates, and these differences can lead to biased estimates of treatment effects. Even traditional covariance analysis adjustments may be inadequate to eliminate this bias. The propensity score, defined as the conditional probability of being treated given the covariates, can be used to balance the covariates in the two groups, and therefore reduce this bias. In order to estimate the propensity score, one must model the distribution of the treatment indicator variable given the observed covariates. Once estimated the propensity score can be used to reduce bias through matching, stratification (subclassification), regression adjustment, or some combination of all three. In this tutorial we discuss the uses of propensity score methods for bias reduction, give references to the literature and illustrate the uses through applied examples. © 1998 John Wiley & Sons, Ltd.
TUTORIAL IN BIOSTATISTICS

PROPENSITY SCORE METHODS FOR BIAS REDUCTION
IN THE COMPARISON OF A TREATMENT TO
A NON-RANDOMIZED CONTROL GROUP

RALPH B. D'AGOSTINO, Jr.*

Department of Public Health Sciences, Section on Biostatistics, Wake Forest University School of Medicine,
Medical Center Boulevard, Winston-Salem, NC 27157-1063, U.S.A.

SUMMARY

In observational studies, investigators have no control over the treatment assignment. The treated and non-treated (that is, control) groups may have large differences on their observed covariates, and these differences can lead to biased estimates of treatment effects. Even traditional covariance analysis adjustments may be inadequate to eliminate this bias. The propensity score, defined as the conditional probability of being treated given the covariates, can be used to balance the covariates in the two groups, and therefore reduce this bias. In order to estimate the propensity score, one must model the distribution of the treatment indicator variable given the observed covariates. Once estimated the propensity score can be used to reduce bias through matching, stratification (subclassification), regression adjustment, or some combination of all three. In this tutorial we discuss the uses of propensity score methods for bias reduction, give references to the literature and illustrate the uses through applied examples. © 1998 John Wiley & Sons, Ltd.
Observational Cohort Studies

Treatment and Control study arms are invariably different in terms of their observed characteristics.

Therefore, quantifying the Treatment effect is complex and requires risk adjustment or application of regression modeling.

Interpretation of Results is **not straightforward**.
TUTORIAL IN BIOSTATISTICS

PROPENSITY SCORE METHODS FOR BIAS REDUCTION
IN THE COMPARISON OF A TREATMENT TO
A NON-RANDOMIZED CONTROL GROUP

RALPH B. D’AGOSTINO, Jr.*

Department of Public Health Sciences, Section on Biostatistics, Wake Forest University School of Medicine,
Medical Center Boulevard, Winston-Salem, NC 27157-1063, U.S.A.

SUMMARY

In observational studies, investigators have no control over the treatment assignment. The treated and non-treated (that is, control) groups may have large differences on their observed covariates, and these differences can lead to biased estimates of treatment effects. Even traditional covariance analysis adjustments may be inadequate to eliminate this bias. The propensity score, defined as the conditional probability of being treated given the covariates, can be used to balance the covariates in the two groups, and therefore reduce this bias. In order to estimate the propensity score, one must model the distribution of the treatment indicator variable given the observed covariates. Once estimated the propensity score can be used to reduce bias through matching, stratification ( subclassification), regression adjustment, or some combination of all three. In this tutorial we discuss the uses of propensity score methods for bias reduction, give references to the literature and illustrate the uses through applied examples. © 1998 John Wiley & Sons, Ltd.
Levels of Evidence

Entirely Dependent on below

Great, But expensive, time consuming, often impractical (or not doable)

Observational, convenient but little control ...

Is there a Magic Bullet

Do Propensity Scores provide this “Magic Bullet”?
TUTORIAL IN BIOSTATISTICS

PROPENSITY SCORE METHODS FOR BIAS REDUCTION
IN THE COMPARISON OF A TREATMENT TO
A NON-RANDOMIZED CONTROL GROUP

RALPH B. D'AGOSTINO, Jr.*

Department of Public Health Sciences, Section on Biostatistics, Wake Forest University School of Medicine,
Medical Center Boulevard, Winston-Salem, NC 27157-1063, U.S.A.

SUMMARY

In observational studies, investigators have no control over the treatment assignment. The treated and non-treated (that is, control) groups may have large differences on their observed covariates, and these differences can lead to biased estimates of treatment effects. Even traditional covariance analysis adjustments may be inadequate to eliminate this bias. The propensity score, defined as the conditional probability of being treated given the covariates, can be used to balance the covariates in the two groups, and therefore reduce this bias. In order to estimate the propensity score, one must model the distribution of the treatment indicator variable given the observed covariates. Once estimated the propensity score can be used to reduce bias through matching, stratification (subclassification), regression adjustment, or some combination of all three. In this tutorial we discuss the uses of propensity score methods for bias reduction, give references to the literature and illustrate the uses through applied examples. © 1998 John Wiley & Sons, Ltd.
Observational Cohort Studies

Treatment and Control study arms are invariably different in terms of their observed characteristics. Therefore, quantifying the Treatment effect is complex and requires risk adjustment or application of regression modeling. Interpretation of Results is not straightforward.

✓ Propensity scores were developed to overcome this difficulty.

✓ The idea is to estimate the likelihood (the propensity score) that a person would have received treatment given certain characteristics (observed covariates).

✓ More formally, the propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates (D’Agostino 1998).
Ways Propensity Scores are Used

1) Propensity Score Stratification: stratify based on PS and compare

2) Propensity Score Adjustment: regression modeling with Treatment and PS

3) Propensity Score Matching:  
   a) 1-to-1; 1-to-many;  
   b) with or without replacement

4) Combinations of 1, 2 and 3: outside today’s scope

5) Inverse Probability Weighting: outside today’s scope
It all comes down to the same “fundamental problem of Causal Inference”

What we want to know is “The Treatment Effect on the Treated”.

But, this is the counterfactual, and we cannot measure outcomes in the same individual, for both treatments, and at the same time.

We need an alternative way to get a credible estimate of the counterfactual outcomes.

It can be shown that one may use the outcomes from the Control arm; provided we have means to handle selection bias.

RCTs are a near perfect way to do that, yet they come at a price.
Examples
Example #1
Right Heart Catheterization (RHC) in ICU

Setting: ICU

Patients: RHC (Treatment) vs. No-RHC (Control)

Outcomes: Death and health care utilization

✓ Groups differed on a large number of risk factors for the outcomes

✓ Expectation: Crude (unadjusted) analysis would be confounded.
Casual Diagram

RHC

- Length of Stay
- Mortality
- Resource Utilization

Treatment: RHC (Yes)
Control: RHC (No)
Age

Diagnosis

Severity

RHC

Mortality

Confounders
Example #2
Effect of Blood Pressure MEDs on Mortality

Cohort: Framingham Heart Study

Subjects: all patients with history of hypertension

Exposure (Treatment): BPMEDS

Outcome: Death

Potential Confounders: AGE, SEX, SBP, DBP, TOTCHOL, BMICAT, CURSMOKE, DIABETES, PREVSTRK, PREVCHD
Confounding by Indication

- Refers to studies assessing a clinical treatment decision (intervention)

- Characterized by a large number of factors that are a) predictors of the outcome and, at the same time, b) influence the treatment decision

“Selection Bias and Need for Risk Adjustment”
Analytic Methods for Risk Adjustment

Stratification

– Simple and Intuitive
– Difficult if too many confounders
Analytic Methods for Risk Adjustment

Regression Modeling

– Outcome Model with terms for exposure and individual confounders.

– Unstable if many confounders
  • 10:1 Rule
Design Methods for Avoiding Confounding

Randomization
– Avoids confounding by known and unknown risk factors.

Matching
– Avoids confounding by matching factors in Cohort Studies
– Difficult to implement for many confounders
Solution

Summarize confounders into a summary score (eg. Propensity Score), then use this score to stratify, adjust or match.
Propensity Score Analysis

✓ Is it a Magic Bullet?

Does it control for all confounders?
   – Known and Unknown?

Does it control for confounding better than traditional methods?

If so, why isn’t it used more often?
Timeline: Use of Propensity Scores

Search Term:
(propensity score OR (propensit* AND (score* OR scoring* OR match* OR adjust*)))
Propensity Score Analysis

Motivation: Control for a large number of confounders

- Combine individual confounders in a summary score (Propensity Score)
- Theory developed > 30 years ago
- Seldom used until recently
  - Particularly used in pharmacoepidemiology, Outcomes and health services research (Confounding by Indication)
Propensity Score Analysis

Propensity Score = Probability of receiving the treatment as a function of the confounders.

– PS Reflects arrows in causal diagrams connecting individual confounders to the treatment.
Going Back to our Examples #1 and #2
Example #1

- Age
- Severity
- Diagnosis

Treatment Decision → RHC → Mortality

Confounders
Example #1

Confounders

- Age
- Severity
- Diagnosis
- Propensity Score
- RHC
- Mortality
Example #2

Treatment Decision

- Age
- Sex
- SBP
- Diabetes

Hypertension Medication

Death
Propensity Score Analysis

Propensity Score = Balancing Score

**Theory:** Patients with similar propensity scores should show similar distributions of confounders within treatment groups (as if treatment had been randomized)

- Conditioning on Propensity Score removes influence of confounders on the intervention
- What remains is random assignment (**plus influence of unknown confounders**)
How To Do IT?
Propensity Score Adjusted Analysis

1. Logistic regression model prediction of:
   a) the Probability of RHC (Example #1) or
   b) the Probability of hypertension Meds (Example #2)
as a function of the respective observed confounders

✓ PS Model does not contain a term for the outcome

2. Regression Model for the Outcome with two factors:
   i) Treatment and ii) PS (or summary score)
Standard Outcome Model vs. PS Adjusted Analysis

- In general, both methods often lead to similar conclusions but not necessarily same values for the adjusted odds ratios or hazard ratios.

- Neither adjust for unmeasured confounders
  - Unless correlated with measured confounders.

- Propensity score analysis may be less sensitive to modeling assumptions.

- Propensity score analysis may be better on cases of many confounders and few outcomes.
Propensity Score Stratification

Discrimination but some overlap of propensity scores in two comparison groups

Analytic plan:

– Create 20 strata defined by sub-ranges of propensity score.
  – (0.00 – 0.05), (0.06 – 0.10), ..., (0.96 – 1.00)

– Check for balance of confounders within strata.

– If balanced, perform stratified analysis.
What to Include in Propensity Score

All confounders

- Leaving a confounder out of PS has same effect as leaving it out of Outcome Model

Many recommend including non-confounders that are risk factors for the outcome

- May improve efficiency.

Don’t include non-confounders that are correlates of the exposure (Treatment)

- May lessen efficiency of analysis.
What about Propensity Matching to simulate RCTs
Improved Survival With Radial Artery Versus Vein Conduits in Coronary Bypass Surgery With Left Internal Thoracic Artery to Left Anterior Descending Artery Grafting
Anoar Zacharias, Robert H. Habib, Thomas A. Schwann, Christopher J. Riordan, Samuel J. Durham and Aamir Shah

Circulation. 2004;109:1489-1496; originally published online March 15, 2004; doi: 10.1161/01.CIR.0000121743.10146.78

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2004 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

Received October 21, 2003; revision received January 9, 2004; accepted January 9, 2004.
From the Division of Cardiovascular Surgery, St Vincent Mercy Medical Center, Toledo (A.Z., R.H.H., T.A.S., C.J.R., S.J.D., A.S.); Division of Cardiovascular Surgery, St Luke’s Hospital, Maumee (A.Z., R.H.H., T.A.S., C.J.R., S.J.D., A.S.); and Departments of Surgery (A.Z., T.A.S., C.J.R., S.J.D., A.S.) and Medicine (R.H.H.), Medical College of Ohio, Toledo, Ohio.

Correspondence to Dr Robert H. Habib, Director, Cardiopulmonary Research, St. Vincent Mercy Medical Center, 2213 Cherry St, ACC Bldg, Suite 309, Toledo, OH 43608. E-mail Robert_Habib@mhsnr.org
© 2004 American Heart Association, Inc.
Circulation is available at http://www.circulationaha.org
DOI: 10.1161/01.CIR.0000121743.10146.78
the LITA-LAD graft. In the present study, we analyzed a large CABG-LITA-LAD experience with the primary aim of investigating whether use of radial versus vein as the second conduit of choice provides improved perioperative and long-term outcomes.
Study

**Population:** Single Institution; N=3161 consecutive Patients (1996-2002) Primary Coronary Artery Bypass Grafting (CABG) surgery

**Groups:** Standard 1-artery CABG (Control) vs. 2-artery CABG (Treatment)

**Outcome:** 6 year Kaplan-Meier Survival – all-cause mortality

**Design:** Propensity Matched Analysis – 1 to 1 Greedy

**Treatment:** LITA graft + Radial Artery Graft

**Control:** LITA graft + Saphenous Vein
Unadjusted Kaplan – Meier Analysis

2 arteries are Better than 1! Confounded?
TABLE 1. Comparison of Overall and Matched Radial Versus Vein Patient Data

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>All Patients</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radial</td>
<td>Vein</td>
</tr>
<tr>
<td>Patients, n</td>
<td>1292</td>
<td>1869</td>
</tr>
<tr>
<td>Male, %</td>
<td>76.9</td>
<td>63.3</td>
</tr>
<tr>
<td>Age, y</td>
<td>61±10</td>
<td>67±10</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>30.5±5.7</td>
<td>29.1±5.4</td>
</tr>
<tr>
<td>BSA, m²</td>
<td>2.1±0.2</td>
<td>2.0±0.2</td>
</tr>
<tr>
<td>Diabetes, %</td>
<td>32.9</td>
<td>37.1</td>
</tr>
<tr>
<td>Renal failure, %</td>
<td>1.2</td>
<td>4.5</td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>81.8</td>
<td>83.2</td>
</tr>
<tr>
<td>Chronic lung disease, %</td>
<td>17.1</td>
<td>23.5</td>
</tr>
<tr>
<td>Peripheral vascular disease, %</td>
<td>10.5</td>
<td>18.6</td>
</tr>
<tr>
<td>Cerebrovascular disease, %</td>
<td>15.6</td>
<td>28.6</td>
</tr>
<tr>
<td>Congestive heart failure, %</td>
<td>8.6</td>
<td>15.8</td>
</tr>
<tr>
<td>Unstable angina, %</td>
<td>54.4</td>
<td>51.3</td>
</tr>
<tr>
<td>Arrhythmia, %</td>
<td>10.1</td>
<td>16.2</td>
</tr>
<tr>
<td>NYHA classification, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I–II</td>
<td>23.0</td>
<td>15.4</td>
</tr>
<tr>
<td>III</td>
<td>47.3</td>
<td>47.5</td>
</tr>
<tr>
<td>IV</td>
<td>29.7</td>
<td>37.1</td>
</tr>
</tbody>
</table>
TABLE 1. Comparison of Overall and Matched Radial Versus Vein Patient Data

<table>
<thead>
<tr>
<th></th>
<th>Radial</th>
<th>Vein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary vessel disease, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triple</td>
<td>75.2</td>
<td>74.2</td>
</tr>
<tr>
<td>Double</td>
<td>22.6</td>
<td>21.9</td>
</tr>
<tr>
<td>Single</td>
<td>2.2</td>
<td>3.6</td>
</tr>
<tr>
<td>Left main disease &gt;50%, %</td>
<td>20.2</td>
<td>22.0</td>
</tr>
<tr>
<td>Ejection fraction, %</td>
<td>50±10</td>
<td>48±11</td>
</tr>
<tr>
<td>Intraoperative data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency, %</td>
<td>3.7</td>
<td>7.1</td>
</tr>
<tr>
<td>Cross clamp, min</td>
<td>48±20</td>
<td>43±17</td>
</tr>
<tr>
<td>Perfusion time, min</td>
<td>79±32</td>
<td>72±29</td>
</tr>
<tr>
<td>Grafts, n</td>
<td>3.4±0.9</td>
<td>3.1±0.8</td>
</tr>
<tr>
<td>Vein</td>
<td>1.0±0.9</td>
<td>2.1±0.8</td>
</tr>
<tr>
<td>Arterial</td>
<td>2.3±0.6</td>
<td>1.0±0.2</td>
</tr>
<tr>
<td>Radial</td>
<td>1.3±0.6</td>
<td>...</td>
</tr>
</tbody>
</table>

Etc ...

BMI indicates body mass index; BSA, body surface area; NYHA, New York Heart Association; IABP, Intra-aortic balloon pump; MI, myocardial infarction; PO, postoperative; and LOS, length of stay. Categorical comparisons were done by χ² test. LOS data were compared by t test.

*P>0.2.
†Power of χ² test was <0.80.
Propensity Matching

Study patients were drawn from observational data, and radial and vein groups were hence characterized by significant demographic and risk factor differences (Table 1). Because such differences substantially influence outcomes, we used propensity matching to overcome such confounding effects, with radial considered treatment. Briefly, the probability that a patient received a radial graft was defined by a propensity score via logistic multivariate modeling applied to all patients. Most risk factors, demographics, and operative variables were entered into the model regardless of significance level. Month of surgery was entered to account for increasing radial use over time. Highly redundant variables were avoided.
PS Distributions (Treatment vs. Control)

Figure 1. Distribution of radial use propensity score in radial (thick lines) and vein (thin lines) patients. Dashed lines represent model fits delineating distinct group distributions; solid lines, propensity score distribution in matched groups.
Propensity Matching Continued

As expected, radial use propensity score distributions derived for radial versus vein patients were distinct (Figure 1). One-to-one propensity matching of radial patients with the closest unique vein match was obtained by a computer algorithm allowing a maximal difference of ±1%. Patient matching was done on the basis of propensity scores only (standard) and by propensity scores in addition to gender and triple-vessel disease status (yes/no) restrictions (restricted).

925 pairs
Were matched

1 to 1 Greedy
PS matching
w/out replacement
Does Propensity Matching Work?
### TABLE 1. Comparison of Overall and Matched Radial Versus Vein Patient Data

<table>
<thead>
<tr>
<th></th>
<th>All Patients</th>
<th></th>
<th>Matched</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radial</td>
<td>Vein</td>
<td>Radial</td>
<td>Vein</td>
</tr>
<tr>
<td>Patients, n</td>
<td>1292</td>
<td>1869</td>
<td>925</td>
<td>925</td>
</tr>
<tr>
<td>Risk factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male, %</td>
<td>76.9</td>
<td>63.3</td>
<td>71.9</td>
<td>71.5</td>
</tr>
<tr>
<td>Age, y</td>
<td>61±10</td>
<td>67±10</td>
<td>63±10</td>
<td>63±10</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>30.5±5.7</td>
<td>29.1±5.4</td>
<td>30.0±5.6</td>
<td>30.1±5.5</td>
</tr>
<tr>
<td>BSA, m²</td>
<td>2.1±0.2</td>
<td>2.0±0.2</td>
<td>2.0±0.2</td>
<td>2.0±0.2</td>
</tr>
<tr>
<td>Diabetes, %</td>
<td>32.9</td>
<td>37.1</td>
<td>34.2</td>
<td>34.3</td>
</tr>
<tr>
<td>Renal failure, %</td>
<td>1.2</td>
<td>4.5</td>
<td>1.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Hypertension, %</td>
<td>81.8</td>
<td>83.2</td>
<td>83.7</td>
<td>80.6</td>
</tr>
<tr>
<td>Chronic lung disease, %</td>
<td>17.1</td>
<td>23.5</td>
<td>18.6</td>
<td>18.3</td>
</tr>
<tr>
<td>Peripheral vascular disease, %</td>
<td>10.5</td>
<td>18.6</td>
<td>12.4</td>
<td>13.3</td>
</tr>
<tr>
<td>Cerebrovascular disease, %</td>
<td>15.6</td>
<td>28.6</td>
<td>18.4</td>
<td>17.6</td>
</tr>
<tr>
<td>Congestive heart failure, %</td>
<td>8.6</td>
<td>15.8</td>
<td>10.3</td>
<td>10.8</td>
</tr>
<tr>
<td>Unstable angina, %</td>
<td>54.4</td>
<td>51.3</td>
<td>54.3</td>
<td>53.9</td>
</tr>
<tr>
<td>Arrhythmia, %</td>
<td>10.1</td>
<td>16.2</td>
<td>12.0</td>
<td>12.5</td>
</tr>
<tr>
<td>NYHA classification, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I–II</td>
<td>23.0</td>
<td>15.4</td>
<td>21.6</td>
<td>19.5</td>
</tr>
<tr>
<td>III</td>
<td>47.3</td>
<td>47.5</td>
<td>47.5</td>
<td>49.9</td>
</tr>
<tr>
<td>IV</td>
<td>29.7</td>
<td>37.1</td>
<td>30.9</td>
<td>30.6</td>
</tr>
<tr>
<td></td>
<td>All Patients</td>
<td></td>
<td>Matched</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------</td>
<td>---------------</td>
<td>----------------</td>
<td>---------------</td>
</tr>
<tr>
<td></td>
<td>Radial</td>
<td>Vein</td>
<td>Radial</td>
<td>Vein</td>
</tr>
<tr>
<td>Coronary vessel disease, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triple</td>
<td>75.2</td>
<td>74.2</td>
<td>75.4</td>
<td>73.5</td>
</tr>
<tr>
<td>Double</td>
<td>22.6</td>
<td>21.9</td>
<td>22.3</td>
<td>23.2</td>
</tr>
<tr>
<td>Single</td>
<td>2.2</td>
<td>3.6</td>
<td>2.4</td>
<td>3.2</td>
</tr>
<tr>
<td>Left main disease &gt;50%, %</td>
<td>20.2</td>
<td>22.0</td>
<td>20.5</td>
<td>20.4</td>
</tr>
<tr>
<td>Ejection fraction, %</td>
<td>50±10</td>
<td>48±11</td>
<td>49±10</td>
<td>49±10</td>
</tr>
<tr>
<td>Intraoperative data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency, %</td>
<td>3.7</td>
<td>7.1</td>
<td>4.3</td>
<td>4.8</td>
</tr>
<tr>
<td>Cross clamp, min</td>
<td>48±20</td>
<td>43±17</td>
<td>46±19</td>
<td>45±19</td>
</tr>
<tr>
<td>Perfusion time, min</td>
<td>79±32</td>
<td>72±29</td>
<td>77±30</td>
<td>76±33</td>
</tr>
<tr>
<td>Grafts, n</td>
<td>3.4±0.9</td>
<td>3.1±0.8</td>
<td>3.3±0.8</td>
<td>3.2±0.2</td>
</tr>
<tr>
<td>Vein</td>
<td>1.0±0.9</td>
<td>2.1±0.8</td>
<td>0.9±0.8</td>
<td>2.2±0.2</td>
</tr>
<tr>
<td>Arterial</td>
<td>2.3±0.6</td>
<td>1.0±0.2</td>
<td>2.3±0.6</td>
<td>1.0±0.2</td>
</tr>
<tr>
<td>Radial</td>
<td>1.3±0.6</td>
<td>...</td>
<td>1.3±0.5</td>
<td>...</td>
</tr>
</tbody>
</table>

Etc ...

BMI indicates body mass index; BSA, body surface area; NYHA, New York Heart Association; IABP, Intra-aortic balloon pump; MI, myocardial infarction; PO, postoperative; and LOS, length of stay. Categorical comparisons were done by $\chi^2$ test. LOS data were compared by $t$ test.

* $P>0.2$.

† Power of $\chi^2$ test was $<0.80$. 
Note,
The primary analysis of the data was simply a Kaplan-Meier comparison Restricted to the matched cohorts.

2 arteries are Better than 1 !

Not Confounded?
Thank you

SHARPIES “Maroun and Karim” will next describe a project they are working on where they are applying what we discussed to great effect.