Propensity Score의 소개와 이용

서울대학교 보건대학원 보건통계학교실

김 호

예제

Hospital Type	Severity of Illness		
	Mild (Mortality Rate %)	Severe (Mortality Rate %)	Total (Mortality Rate %)
World-Class Hospitals	3/327 (1%)	41/678 (6%)	44/1,005 (4.4%)
Local Hospitals	8/258 (3%)	3/33 (9%)	11/291 (3.8%)
Total	11/585 (2%)	44/711 (3%)	

- sever 환자가 큰 병원으로 간다. 사망률은 큰 병원에서 더 높다.
- Treatment selection이 severity와 독립이 아니다.
- 병원비교에 있어서 severity는 중요한 혼란변수.
- Typical “Apples to Oranges” comparisons

예제

Hospital Type	Severity of Illness		
	Mild (Mortality Rate %)	Severe (Mortality Rate %)	Total (Mortality Rate %)
World-Class Hospitals	3/327 (1%)	41/678 (6%)	44/1,005 (4.4%)
Local Hospitals	8/258 (3%)	3/33 (9%)	11/291 (3.8%)
Total	11/585 (2%)	44/711 (3%)	

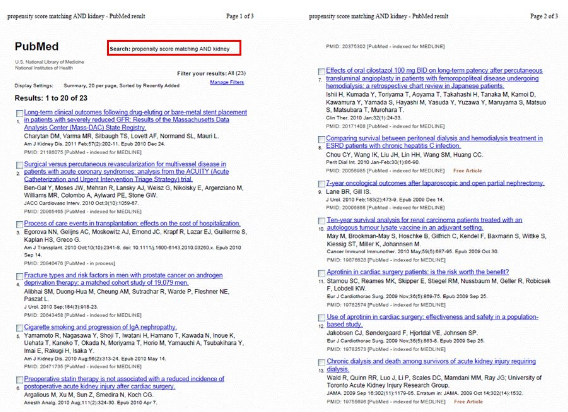
- sever 환자가 큰 병원으로 간다. 사망률은 큰 병원에서 더 높다.
- Treatment selection이 severity와 독립이 아니다.
- 병원비교에 있어서 severity는 중요한 혼란변수.
- Typical “Apples to Oranges” comparisons

Limitations of Random Assignment

- Large RCTs take a long time and great cost to generate answers—analysis of existing data may more timely, yet acceptably accurate
- RCTs are not feasible when variables cannot be manipulated—e.g., some events in child welfare are driven by legal mandates
- Prior analysis of the need for withholding treatment should be done before RCTs are deemed necessary.

Propensity Score Analysis

- Motivation
  - Control for a large number of confounders by a adjusting for a single summary variable
- Propensity Score
  - Predicts the probability of receiving the intervention as a function of the confounders



propensity score matching AND kidney - PubMed result Page 1 of 3

PubMed

Search: propensity score matching AND kidney

U.S. National Library of Medicine
National Institutes of Health

Filter your results: All (23)

Manage Filters

Display Settings: Summary, 20 per page, Sorted by Recently Added

Results: 1 to 20 of 23

1. Long-term clinical outcomes following drug-eluting or bare-metal stent placement in patients with severely reduced GFR: Results of the Massachusetts Data Analysis Center (Mass-DAC) State Registry.
Charytan DM, Varma MR, Silbaugh TS, Lovett AF, Normand SL, Mauri L.
Am J Kidney Dis. 2011 Feb;57(2):202-11. Epub 2010 Dec 24.
PMID: 21186075 [PubMed - indexed for MEDLINE]

2. Surgical versus percutaneous revascularization for multivessel disease in patients with acute coronary syndromes: analysis from the ACUITY (Acute Catheterization and Urgent Intervention Triage Strategy) trial.
Ben-Gal Y, Moses JW, Mehran R, Lansky AJ, Weisz G, Nikolsky E, Argenziano M, Williams MR, Colombo A, Aylward PE, Stone GW.
JACC Cardiovasc Interv. 2010 Oct;3(10):1059-67.
PMID: 20965465 [PubMed - indexed for MEDLINE]

3. Process of care events in transplantation: effects on the cost of hospitalization.
Egorova NN, Gelijns AC, Moskowitz AJ, Emond JC, Krapf R, Lazar EJ, Guillerme S, Kaplan HS, Greco G.
Am J Transplant. 2010 Oct;10(10):2341-8. doi: 10.1111/j.1600-6143.2010.03260.x. Epub 2010 Sep 14.
PMID: 20840476 [PubMed - in process]

4. Fracture types and risk factors in men with prostate cancer on androgen deprivation therapy: a matched cohort study of 19,079 men.
Alibhai SM, Duong-Hua M, Cheung AM, Sutradhar R, Warde P, Fleshner NE, Paszat L.
J Urol. 2010 Sep;184(3):918-23.
PMID: 20643458 [PubMed - indexed for MEDLINE]

5. Cigarette smoking and progression of IgA nephropathy.
Yamamoto R, Nagasawa Y, Shoji T, Iwatani H, Hamano T, Kawada N, Inoue K, Uehata T, Kaneko T, Okada N, Moriyama T, Horio M, Yamauchi A, Tsubakihara Y, Imai E, Rakugi H, Isaka Y.
Am J Kidney Dis. 2010 Aug;56(2):313-24. Epub 2010 May 14.
PMID: 20471735 [PubMed - indexed for MEDLINE]

6. Preoperative statin therapy is not associated with a reduced incidence of postoperative acute kidney injury after cardiac surgery.
Argalious M, Xu M, Sun Z, Smedira N, Koch CG.
Anesth Analg. 2010 Aug;111(2):324-30. Epub 2010 Apr 7.

propensity score matching AND kidney - PubMed result Page 2 of 3

PMID: 20375302 [PubMed - indexed for MEDLINE]

7. Effects of oral cilostazol 100 mg BID on long-term patency after percutaneous transluminal angioplasty in patients with femoropopliteal disease undergoing hemodialysis: a retrospective chart review in Japanese patients.
Ishii H, Kumada Y, Toriyama T, Aoyama T, Takahashi H, Tanaka M, Kamoi D, Kawamura Y, Yamada S, Hayashi M, Yasuda Y, Yuzawa Y, Maruyama S, Matsuo S, Matsubara T, Murohara T.
Clin Ther. 2010 Jan;32(1):24-33.
PMID: 20171408 [PubMed - indexed for MEDLINE]

8. Comparing survival between peritoneal dialysis and hemodialysis treatment in ESRD patients with chronic hepatitis C infection.
Chou CY, Wang IK, Liu JH, Lin HH, Wang SM, Huang CC.
Perit Dial Int. 2010 Jan-Feb;30(1):86-90.
PMID: 20056985 [PubMed - indexed for MEDLINE] Free Article

9. 7-year oncological outcomes after laparoscopic and open partial nephrectomy.
Lane BR, Gill IS.
J Urol. 2010 Feb;183(2):473-9. Epub 2009 Dec 14.
PMID: 20006866 [PubMed - indexed for MEDLINE]

10. Ten-year survival analysis for renal carcinoma patients treated with an autologous tumour lysate vaccine in an adjuvant setting.
May M, Brookman-May S, Hoschke B, Gilfrich C, Kendel F, Baxmann S, Wittke S, Kiessig ST, Miller K, Johannsen M.
Cancer Immunol Immunother. 2010 May;59(5):687-95. Epub 2009 Oct 30.
PMID: 19876628 [PubMed - indexed for MEDLINE]

11. Aprotinin in cardiac surgery patients: is the risk worth the benefit?
Stamou SC, Reames MK, Skipper E, Stiegel RM, Nussbaum M, Geller R, Robicsek F, Lobdell KW.
Eur J Cardiothorac Surg. 2009 Nov;36(5):869-75. Epub 2009 Sep 25.
PMID: 19782574 [PubMed - indexed for MEDLINE]

12. Use of aprotinin in cardiac surgery: effectiveness and safety in a population-based study.
Jakobsen CJ, Søndergaard F, Hjortdal VE, Johnsen SP.
Eur J Cardiothorac Surg. 2009 Nov;36(5):863-8. Epub 2009 Sep 25.
PMID: 19782573 [PubMed - indexed for MEDLINE]

13. Chronic dialysis and death among survivors of acute kidney injury requiring dialysis.
Wald R, Quinn RR, Luo J, Li P, Scales DC, Mamdani MM, Ray JG; University of Toronto Acute Kidney Injury Research Group.
JAMA. 2009 Sep 16;302(11):1179-85. Erratum in: JAMA. 2009 Oct 14;302(14):1532.
PMID: 19755696 [PubMed - indexed for MEDLINE] Free Article

What are propensity scores?

- ✓ is the predicted probability of an outcome (SUGI 29;165)
- ✓ is generally defined as the conditional probability of assignment to a particular treatment given a vector of observed covariates (Rosenbaum and Rubin 1983)
- ✓ are the predicted probabilities from a logistic model which models the probabilities of being in the various levels of the predictor of primary interest as a function of a set of secondary variables.

Propensity Scores

- Suppose we want to look at a response (expression of a certain outcome) if a patient received a treatment, compared to the response if the patient did not receive the treatment.
- Dataset we wish to have

Patient	Outcome if received treatment	Outcome if received no treatment	Treatment impact
A	12	8	4
B	7	4	3
C	7	3	4
D	12	9	3

Propensity Scores

- What is in reality?

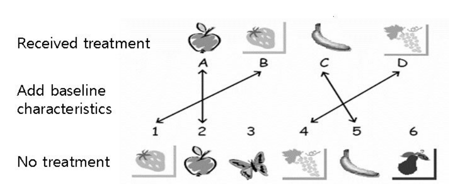
Patient	Outcome if received treatment	Outcome if received no treatment	Treatment impact
A	12	?	?
B	7	?	?
C	?	3	?
D	?	9	?

Propensity Scores

- The question is …
  - If we have a patient who got the treatment in the dataset, how can we specify what that patient's response would have been if they had not gotten the treatment?
  - The treatment effect is the difference between the two responses (treatment – no treatment)
  - Rubin Causal Model – potential outcomes framework

Propensity Scores

- Comparing apples to apples



Propensity Scores

- Propensity scores
  - We have a lot of information on each potential subject.
    - Demographics
    - Geographic/socioeconomic information
  - We can estimate the probability that a subject with these background characteristics will receive the treatment.

$0 \le PS \le 1$

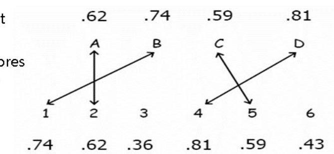
Propensity Scores

- Comparing apples to apples

Received treatment

Add propensity scores based on baseline characteristics

No treatment



Propensity Scores

- The new dataset, simply

Patient	Propensity to get treatment	Outcome if received treatment	Outcome if received no treatment	Treatment impact
A	0.8	12	9	3
B	0.5	7	3	4
C	0.5	7	3	4
D	0.8	12	9	3

- Match A to D and B to C and plug in resulting estimates for blanks…

General Procedure

Run Logistic Regression:

- Dependent variable: Y=1, if participate; Y = 0, otherwise.
- Choose appropriate conditioning (instrumental) variables.
- Obtain propensity score: predicted probability (p) or log[(1-p)/p].

→

Match each Participant to One or More Nonparticipants on Propensity Score

- Nearest neighbor matching
- Caliper matching
- Mahalanobis Metric matching in conjunction with PSM
- stratification (subclassification)
- Difference-in-differences matching [Kernel & local linear weights (Heckman)]

→

Multivariate analysis based on new sample

Introduction to Propensity Score Matching
By Shenyang Guo, Ph.D. (sguo@email.unc.edu)
School of Social Work. University of North Carolina at Chapel Hill

Calculation of Propensity Score

- Propensity scores are generally calculated using one of two methods:
  1. **Logistic regression** : This is the most commonly used method for estimating propensity scores.
  2. Classification and Regression Tree Analysis(CART): It is a model used to predict the probability that an event occurs.

•Requirement
1. classification ability : C statistic
   - Treatment여부와 PS간에 ROC curve를 그렸을 때 AUC
   - 0.75이상이면 만족
2. calibration ability: Hosmer-Lameshow test
   - LR model에서 Goodness-of-fit test
   - p가 significant한 level이 아니면 만족.

Classification and Regression tree analysis (CART)

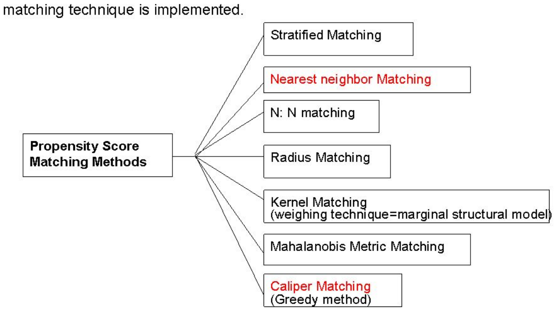
- This is ***a non-parametric decision tree method*** that can efficiently partition populations into homogenous subgroups (Lemon, Freidmann, Rakowski, 2003).
- It is not as widely used as logistic regression for estimating propensity scores (because it is complex and more suitable for use by those with a statistical background)

Adjustment for the estimated propensity scores

- Adjustment for the estimated propensity scores is accomplished using one or a combination of the four main methods.
  - (1) Stratification,
  - **(2) Matching,**
  - (3) Covariate/Regression adjustment, and
  - (4) Weighting
- Once the propensity scores are estimated, these methods can be used to estimate the treatment effect after adjusting for differences between the treatment groups.
- Both stratification and matching are used to adjust for the covariate **before** calculating the treatment effect. In contrast, regression adjustment is used **while** determining the treatment effect.

Propensity Score Matching Methods

•Once researchers obtain an estimated propensity score, an appropriate matching technique is implemented.



Nearest Neighbor and Caliper Matching

- Nearest neighbor:
  - Randomly order the participants and nonparticipants, then select the first participant and find the nonparticipantwith closest propensity score
- Caliper
  - define a common-support region (e.g., .01 to .00001), and randomly select one nonparticipantthat matches on the propensity score with the participant.
  - SAS macro "**GREEDY**" does this.
- 1-to-1 Nearest neighbor within caliper (common practice)
- 1-to-n Nearest neighbor within caliper

Introduction to Propensity Score Matching
By Shenyang Guo, Ph.D. (sguo@email.unc.edu)
School of Social Work. University of North Carolina at Chapel Hill

Stratification

One of several methods developed for missing data imputation

1. Group sample into five categories based on propensity score (quintiles).
2. Within each quintile, there are *r* participants and *n* nonparticipants. Use "approximate Bayesian bootstrap" method to conduct matching or resampling.

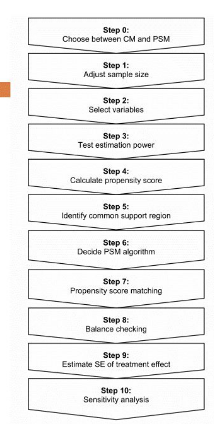
Multivariate Analysis

We could perform any kind of multivariate analysis we originally wished to perform on the unmatched data. These analyses may include:

- multiple regression
- generalized linear model
- survival analysis
- structural equation modeling with multiple-group comparison, and
- hierarchical linear modeling (HLM)

Introduction to Propensity Score Matching
By Shenyang Guo, Ph.D. (sguo@email.unc.edu)
School of Social Work. University of North Carolina at Chapel Hill

PSM implementation steps

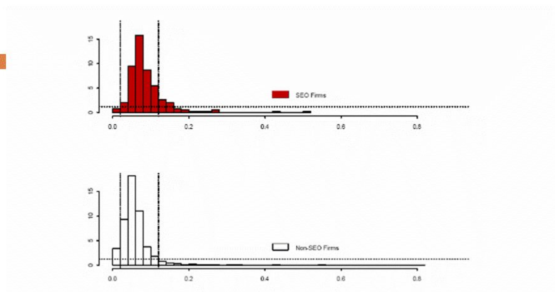


CM: Covariate matching

PSM: Propensity score matching

SE: Standard error

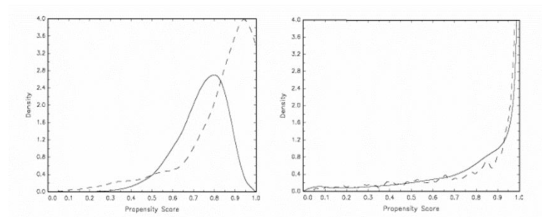
Yang et al.



Common Support Region and Trimming Level (Li and Zhao, 2006)
(The horizontal broken lines in both treatment and control distributions represent the cutoff level q; the left and right vertical lines define common support region points.)

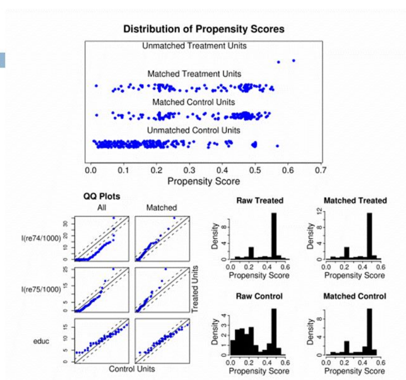
Yang et al.

Balance Checking



Non-balanced Matching **Balanced Matching**
(Dashed curve is treatment, solid curve is control, from Shaikh and coworkers, 2005,)

Yang et al.



Ho et al.

Table 2: 1:1 Matched Population

	Treatment N (%)	No Treatment N (%)	p-value
Total Patients	**3,905**	**3,905**	
Age Mean ± sd	67.1± 12.3	66.8 ± 13.7	0.5596
Female	1,481 (37.9)	1,468 (37.6)	0.7616
White Race	3,283 (84.1)	3,259 (83.5)	0.4615
ECG Obtained	300 (7.9)	278 (7.2)	0.2820
Sx to Door(hr) Mean ± sd	2.2 ± 2.0	2.2 ± 2.2	0.1327
Hx Angina	483 (12.4)	479 (12.3)	0.8905
Hx MI	872 (22.3)	888 (22.7)	0.6648
Hx CHF	259 (6.6)	243 (6.2)	0.4604
Hx CABG	340 (8.7)	331 (8.5)	0.7163
Hx PTCA	511 (13.1)	506 (13.0)	0.8665
Hx Diabetes	931 (23.8)	931 (23.8)	0.7505
Hx Smoking	1,150 (29.4)	1,122 (28.7)	0.4855
Hx Hyperchol.	1,116 (28.6)	1,110 (28.4)	0.8805
Hx CAD	963 (24.7)	950 (24.3)	0.7323
Hx Hyperten.	2,340 (59.9)	2,388 (61.2)	0.2665
Hx Stroke	1,475 (37.8)	1,454 (37.2)	0.6236
Hx Renal	118 (5.9)	125 (6.1)	0.8311
Hx COPD	243 (12.2)	231 (11.2)	0.3439
Chest Pain	3,401 (89.1)	3,411 (88.6)	0.4846
Killip Class IV	183 (4.7)	187 (4.8)	0.8313
Systolic BP Mean ± sd	139.8 ± 35.8	141.0 ± 37.0	0.8313
Diastolic BP Mean ± sd	80.7 ± 20.5	81.2 ± 20.9	0.5005
Pulse Mean ± sd	81.5 ± 23.3	82.0 ± 23.0	0.1378
Propensity Score Mean ± sd min, max	.3650± .1800 .0064, .8833	.3649± .1801 .0064, .8780	

Table 3: 1:2 Matched Population

	Treatment N (%)	No Treatment N (%)	p-value
Total Patients	**2,164**	**4,328**	
Age Mean ± sd	71.7±10.9	72.5±11.7	0.0001
Female	976 (45.1)	1,946 (44.96)	0.9157
White Race	1,833 (84.7)	3,692 (85.3)	0.5216
ECG Obtained	104 (4.8)	209 (4.8)	0.9673
Sx to Door(hr) Mean ± sd	2.6 ± 2.2	2.6 ± 2.4	0.1713
Hx Angina	312 (14.4)	616 (14.2)	0.8410
Hx MI	578 (26.7)	1,138 (26.3)	0.7202
Hx CHF	237 (11.0)	472 (10.9)	0.9551
Hx CABG	281 (13.0)	540 (12.5)	0.5613
Hx PTCA	213 (9.8)	369 (8.5)	0.0799
Hx Diabetes	634 (29.3)	1,216 (28.1)	0.3120
Hx Smoking	470 (21.7)	874 (20.2)	0.1528
Hx Hyperchol.	518 (23.9)	956 (22.1)	0.0938
Hx CAD	462 (21.4)	892 (20.6)	0.4895
Hx Hyperten.	1,350 (72.4)	2,691 (62.2)	0.8706
Hx Stroke	766 (35.4)	1,685 (38.9)	0.0056
Hx Renal	106 (4.9)	201 (4.6)	0.6492
Hx COPD	155 (7.2)	337 (7.8)	0.3706
Chest Pain	1,683 (77.8)	3,426 (79.2)	0.1984
Killip Class IV	107 (4.9)	175 (4.0)	0.0931
Systolic BP Mean ± sd	142.4 ± 36.9	142.1 ± 36.5	0.2567
Diastolic BP Mean ± sd	80.4 ± 20.3	80.1 ± 19.8	0.5124
Pulse Mean ± sd	87.2 ± 25.0	86.8 ± 23.6	0.2776
Propensity Score Mean ± sd min, max	.2330± .1024 .0064, .4477	.2329± .1023 .0064, .4485	

Parson et al.

Example

- Aspirin Use and Mortality
  - 6174 consecutive adults undergoing stress echocardiography for evaluation of known suspected coronary disease.
  - 2310 (37%) were taking aspirin (treatment).
  - Main outcome: all-cause mortality
  - Median follow-up: 3.1 years
  - Univariate analysis: 4.5% of aspirin patients died, and 4.5% of non-aspirin patients died …
  - Unadjusted hazard ratio: 1.08 (0.85, 1.39)

Example

- Propensity Score Model for Aspirin Use
  - Logistic regression predicting aspirin use
  - 31 covariates included in the model
    - Demographics, clinical history, medication use
    - Cardiovascular assessment and exercise capacity
  - Estimated propensity scores for aspirin use range from 0.03 to 0.98.
  - Does the propensity score model work?
  - Are the covariates balanced?

Example

- Baseline Characteristics by Aspirin Use (in %) (before Matching)

Variable	Aspirin (n = 2310)	No Aspirin (n = 3864)	P value
Men	77.0	56.1	< .001
Clinical history: diabetes	16.8	11.2	< .001
hypertension	53.0	40.6	< .001
prior coronary artery disease	69.7	20.1	< .001
congestive heart failure	5.5	4.6	.12
Medication use: Beta-blocker	35.1	14.2	< .001
ACE inhibitor	13.0	11.4	< .001

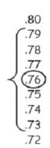
- Baseline characteristics appear very dissimilar: 25 of 31 covariates have p < .001, 28 of 31 have p < .05.
- Aspirin user covariates indicate higher mortality risk.

Example

- Matching with Propensity Scores
  - Fore each patient, we have a propensity score.
  - Match to the non-user with closest propensity score (within some limit or "calipers")
  - Eliminate both patients from pool, and repeat until you cannot find an acceptable match.
    - Could match a non-user with propensity score inside "calipers" who matches exactly on characteristic **X**, or …
    - Match non-user with propensity score inside "calipers" and smallest "distance" on some pre-specified covariates.

Example

- Matching on Gender within PS calipers
  - Shuffle "treatment" patients, and select one.
  - Find all "non-treated" with PS inside calipers (here we set calipers at treated PS ± 0.03)
  - Match patient within calipers of same gender.
  - Repeat until no more matches are possible



Patient	Exposure	PS	Gender
A	**Treated**	**.76**	**Male**
B	Not Treated	.77	Female
C	Not Treated	.74	Male
D	Not Treated	.80	Male

Example

- How Were the Aspirin Subjects Matched?
  - Tried to match each aspirin user to a unique non-user with a PS identical to 5 digits.
  - If not possible, proceeded to a 4-digit match, then 3-digit, 2-digit, and finally 1-digit match (i.e., propensity scores within .099).
  - Result: matches for 1351(58%) of the 2310 aspirin patients to 1351 unique non-users.

Example

- Baseline Characteristics by Aspirin Use (%) (after Matching)

Variable	Aspirin (n = 1351)	No Aspirin (n = 1351)	P value
Men	70.4	72.1	.33
Clinical history: diabetes	15.0	15.3	.83
hypertension	50.3	51.7	.46
prior coronary artery disease	48.3	48.8	.79
congestive heart failure	5.8	6.6	.43
Medication use: Beta-blocker	26.1	26.5	.79
ACE inhibitor	15.5	15.8	.79

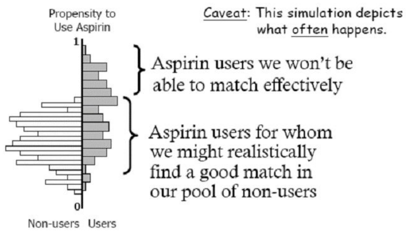
- Baseline characteristics similar in matched users and non-users.
- 30 of 31 covariates show NS difference between matched users and non-users. [Peak exercise capacity for men is p = .01]

Example

- Which Aspirin Users Get Matched?
  - Generally, characteristics of unmatched aspirin users tend to indicate high propensity scores.
    - Overall 37% of patients were taking aspirin.
    - The rate was much higher in some populations… 67% of Prior CAD patients were taking aspirin.
    - So, prior CAD patients had higher propensity for aspirin.
    - 99.8% of unmatched aspirin users had prior CAD.
  - Likely that unmatched users tended towards larger propensity scores than matched users.

Example

- Who is Getting Matched Here? Where Do the Propensity Scores Overlap?



Example

- Matching with Propensity Scores
  - 1351 aspirin subjects matched well to non-aspirin subjects – big improvement in covariate balance. Matched group looks like an RCT …
  - Matching still incomplete, but results on PS matched group mirrored the results for the covariate-adjusted group as a whole …
  - Resulting matched pairs analyzed using standard statistical methods, e.g. Kaplan-Meier, Cox PH models.

Example

- Estimating the Hazard Ratios

Approach	n	Hazard Ratio	95% CI
Full sample, no adjustment	6174	**1.08**	(.85, 1.39)
Full sample with no PS, adjusted for all covariates	6174	**0.67**	(.51, .87)
PS-Matched sample	2702	**0.53**	(.38, .74)
PS-Matched, adjusted for PS and all covariates	2702	**0.56**	(.40, .78)

- During follow-up 153 (6%) of the 2702 propensity score-matched patients died.
- Aspirin use was associated with a lower risk of death in matched group (4% vs. 8%, p = .002).

Example

- Aspirin Conclusions / Caveats
  - Patients included in this study may be a more representative sample of "real world" patients than an RCT would provide.
  - PS matching is still not randomization: can only account for the factors measured, and only as well as the instruments can measure them.
  - No information on aspirin dose, aspirin allergy, or duration of treatment, or on medication adjustments.

Discussion

- Advantages in PS method
  - It is useful when adjusting for a large number of confounders and small number of events per variable.
- Limitations in PS method
  - PSM cannot adjust for unobserved covariates,
  - PSM works better in larger samples, and
  - Group overlap is substantial in PSM.
  - PS model should be correct.
- PS method can be used for ***missing data analysis*** under missing at random(MAR) assumption.

What should always be done in an observational study and often is not?

- Collect data so as to be able to model selection
- Demonstrate selection bias – need for PS
- Ensure covariate overlap for comparability
- Evaluate covariate balance after PS application
- Specify relevant post-adjustment population carefully
- Model or estimate treatment effect in light of PS (adjustment / matching / stratification)
- Estimate sensitivity of results to potential hidden bias.

예시 (1)

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 APRIL 24, 2008 VOL. 358 NO. 17

Stents versus Coronary-Artery Bypass Grafting for Left Main Coronary Artery Disease

Ki Bae Seung, M.D., Duk-Woo Park, M.D., Young-Hak Kim, M.D., Seung-Whan Lee, M.D., Cheol Whan Lee, M.D., Myeong-Ki Hong, M.D., Seong-Wook Park, M.D., Sung-Cheol Yun, Ph.D., Hyeon-Cheol Gwon, M.D., Myung-Ho Jeong, M.D., Yangsoo Jang, M.D., Hyo-Soo Kim, M.D., Pum Joon Kim, M.D., In-Whan Seong, M.D., Hun Sik Park, M.D., Taehoon Ahn, M.D., In-Ho Chae, M.D., Seung-Jea Tahk, M.D., Wook-Sung Chung, M.D., and Seung-Jung Park, M.D.

Table 1. Baseline Characteristics of the Patients.*

Variable	Stent Group (N=1102)	CABG Group (N=1138)	P Value
Demographic characteristics			
Age (yr)			<0.001
Median	62	64	
Interquartile range	52–70	57–70	
Male sex (%)	70.7	72.9	0.24
Cardiac or coexisting conditions			
Diabetes mellitus (% of patients)			
Any diabetes	29.7	34.7	0.01
Insulin-dependent	6.8	8.2	0.22
Hypertension (% of patients)	49.5	49.4	0.94
Hyperlipidemia (% of patients)	28.5	32.6	0.04
Current smoker (% of patients)	25.6	29.8	0.03
Previous coronary angioplasty (% of patients)	18.1	11.0	<0.001
Previous myocardial infarction (% of patients)	8.1	11.6	0.005
Previous congestive heart failure (% of patients)	2.5	3.3	0.21
Chronic obstructive pulmonary disease (% of patients)	2.0	2.0	0.97
Cerebrovascular disease (% of patients)	7.1	7.3	0.84
Peripheral vascular disease (% of patients)	1.5	5.4	<0.001
Renal failure (% of patients)	2.7	3.0	0.71
Ejection fraction (%)			<0.001
Median	62	60	
Interquartile range	57–67	52–66	
Electrocardiographic findings (% of patients)			0.53
Sinus rhythm	97.8	97.1	
Atrial fibrillation	2.0	2.7	
Other	0.2	0.2	
Clinical indication (% of patients)			<0.001
Silent ischemia	3.0	2.2	
Chronic stable angina	32.0	19.9	
Unstable angina	55.2	68.1	
Non–ST-elevation myocardial infarction	9.8	9.8	
Angiographic characteristics			
Involved location (% of patients)			0.04
Ostium, midshaft, or both	50.6	46.2	
Distal bifurcation	49.4	53.8	
Extent of diseased vessel (% of patients)			<0.001
Left main only	25.2	6.2	
Left main plus single-vessel disease	24.0	10.5	
Left main plus double-vessel disease	26.0	26.3	
Left main plus triple-vessel disease	24.8	57.0	
Right coronary artery disease (% of patients)	35.9	70.7	<0.001
Restenotic lesion (% of patients)	2.9	1.2	0.005

Methods of Propensity-Score Matching (1)

- The propensity scores were estimated without regard to outcome variables, using **multiple logistic-regression analysis.**
  - All prespecified covariates, which are listed in Table 1 of the article, were included in the final models for treatment with PCI versus CABG along with pertinent second-order interactions using step-wise selection procedures.
  - A propensity score was then calculated from the logistic equation for each patient.
  - Also, for each concurrent comparison (Wave 1 and Wave 2), a separate propensity score for PCI versus CABG was generated using the same method for the overall cohort and was incorporated for each analysis.
- The predictive ability of each propensity-score model was assessed by means of the C statistic (0.86 for the entire cohort, 0.88 for Wave 1, and 0.86 for Wave 2), indicating good discrimination between two treatments.

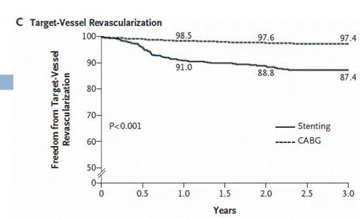
Methods of Propensity-Score Matching (2)

- Using the "**Greedy 5→1 digit match algorithm**", we created **propensity-score-matched pairs** without replacement (a 1:1 match).
  - Specifically, we sought to match each patient with PCI to one with CABG who had a propensity score that was identical to 5 digits.
  - If this could not be done, the algorithm then proceeded sequentially to the next highest digit match (a 4-, 3-, 2-, or 1-digit) on propensity score to make "next-best" matches, in a hierarchical sequence until no more matches could be made.
  - If a subject who received PCI could not be matched to any subject who received CABG on the first digit of the propensity score, that subject with PCI was discarded from the matched analysis.
  - Once a match was made, previous matches were not reconsidered before making the next match.
  - After all of the propensity-scores-matches were performed, we assessed the balance in baseline covariates between the two intervention groups with the paired *t*- test or the Wilcoxon signed rank test for continuous variables, and the McNemar's test or marginal homogeneity test for categorical variables.

Table 2. Baseline Characteristics of the Propensity-Matched Patients.*

Variable	Overall Cohort: Stents (N=542)	Overall Cohort: CABG (N=542)	Overall Cohort: P Value§	Wave 1†: Bare-Metal Stents (N=207)	Wave 1†: CABG (N=207)	Wave 1†: P Value§	Wave 2‡: Drug-Eluting Stents (N=396)	Wave 2‡: CABG (N=396)	Wave 2‡: P Value§
Age (yr)			0.41			0.57			0.23
Median	64	64		61	61		66	66	
Interquartile range	56–71	56–70		51–69	53–67		57-72	58-70	
Male sex (% of patients)	71.6	71.2	0.95	72.0	71.0	0.91	71.5	71.7	0.99
Diabetes mellitus (% of patients)									
Any diabetes	32.7	33.0	0.95	26.1	26.6	0.99	36.1	36.9	0.86
Insulin-dependent	7.6	7.9	0.91	4.8	5.3	0.99	10.1	10.9	0.77
Hypertension (% of patients)	49.4	50.0	0.90	44.9	45.4	0.99	52.3	53.0	0.81
Hyperlipidemia (% of patients)	29.3	30.1	0.84	27.1	27.1	0.99	32.6	33.6	0.81
Current smoker (% of patients)	27.7	27.1	0.89	28.5	28.0	0.99	26.3	25.5	0.87
Previous coronary angioplasty (% of patients)	14.8	15.1	0.93	14.0	14.5	0.99	15.4	15.4	0.99
Previous myocardial infarction (% of patients)	9.0	10.0	0.68	9.7	10.6	0.87	8.8	9.3	0.90
Previous congestive heart failure (% of patients)	2.8	3.0	0.99	2.4	2.9	0.99	3.0	3.3	0.99
Chronic obstructive pulmonary disease (% of patients)	2.6	2.2	0.85	2.4	1.9	0.99	2.8	2.5	0.99
Cerebrovascular disease (% of patients)	7.4	6.6	0.72	6.8	6.3	0.99	8.1	7.3	0.76
Peripheral vascular disease (% of patients)	2.0	2.0	0.99	1.0	1.0	0.99	2.5	3.3	0.79
Renal failure (% of patients)	3.7	3.9	0.99	1.9	2.4	0.99	5.3	4.8	0.83
Ejection fraction (%)			0.62			0.67			0.53
Median	61	61		61	61		60	60	
Interquartile range	54–66	55–66		57–67	56–66		55-66	56-66	
Electrocardiographic findings (% of patients)			0.80			0.99			0.62
Sinus rhythm	97.6	96.7		97.6	97.1		97.7	96.5	
Atrial fibrillation	2.4	3.1		2.4	2.9		2.3	3.0	
Other	0.0	0.2		0	0		0	0.5	
Clinical indication (% of patients)			0.97			0.78			0.41
Silent ischemia	2.8	2.6		2.9	3.4		2.3	2.8	
Chronic stable angina	29.2	28.4		16.4	16.4		30.1	28.8	
Unstable angina	57.4	57.9		69.6	69.6		57.8	57.8	
Non–ST-elevation myocardial infarction	10.7	11.1		11.1	10.6		9.8	10.6	
Involved location (% of patients)			0.90			0.99			0.93
Ostium, midshaft, or both	48.3	47.8		61.8	61.4		39.4	38.9	
Distal bifurcation	51.7	52.2		38.2	38.6		60.6	61.1	
Extent of diseased vessel (% of patients)			0.81			0.99			0.63
Left main only	11.8	11.1		21.3	21.3		5.8	5.8	
Left main plus single-vessel disease	17.0	16.2		29.0	29.0		12.4	11.6	
Left main plus double-vessel disease	31.7	33.9		33.8	33.8		29.0	29.5	
Left main plus triple-vessel disease	39.5	38.7		15.9	15.9	0.99	52.8	53.0	
Right coronary artery disease (% of patients)	53.7	53.7	0.99	29.5	29.5	0.99	65.9	66.9	0.64
Restenotic lesion (% of patients)	1.8	1.8	0.99	1.9	2.4	0.99	1.8	1.3	0.77

* CABG denotes coronary-artery bypass grafting. Percentages may not total 100 because of rounding.
† Wave 1 shows comparisons between bare-metal stents and CABG.
‡ Wave 2 shows comparisons between drug-eluting stents and CABG.
§ P values are based on the paired t-test or the Wilcoxon signed-rank test for continuous variables and on the McNemar's test or marginal homogeneity test for categorical variables.



No. at Risk	0	1.0	2.0	3.0
Stenting	542	471	331	193
CABG	542	503	408	305

Table 3. Hazard Ratios for Clinical Outcomes after Stenting as Compared with after CABG among Propensity-Matched Patients.*

Outcome	Overall Cohort (N=542 pairs): Hazard Ratio (95% CI)	Overall Cohort: P Value	Wave 1 (N=207 pairs): Hazard Ratio (95% CI)	Wave 1: P Value	Wave 2 (N=396 pairs): Hazard Ratio (95% CI)	Wave 2: P Value
Death	1.18 (0.77–1.80)	0.45	1.04 (0.59–1.83)	0.90	1.36 (0.80–2.30)	0.26
Composite outcome of death, Q-wave myocardial infarction, or stroke	1.10 (0.75–1.62)	0.61	0.86 (0.50–1.49)	0.59	1.40 (0.88–2.22)	0.15
Target-vessel revascularization	4.76 (2.80–8.11)	<0.001	10.70 (3.80–29.90)	<0.001	5.96 (2.51–14.10)	<0.001

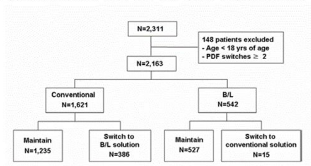
* CABG denotes coronary-artery bypass grafting. Wave 1 shows comparisons between bare-metal stents and CABG, and Wave 2 shows comparisons between drug-eluting stents and CABG. Hazard ratios are for the stenting group as compared with the CABG group.

예시 (2)

Mortality and Technique Failure in Peritoneal Dialysis Patients Using Advanced Peritoneal Dialysis Solutions

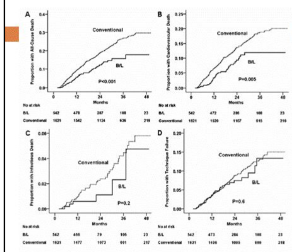
Seung Hyeok Han, MD, PhD,[1,2] Song Vogue Ahn, MD,[3] Jee Young Yun, CRN,[4] Anders Tranaeus, MD, PhD,[5] and Dae-Suk Han, MD, PhD[1]

- conventional lactate-buffered glucose based (Dianeal®) vs. B/L buffer [biocompatible physiological-pH bicarbonate(25 mmol/L)/lactate(15 mmol/L) buffer; Physioneal®]
- 2,163 patients starting peritoneal dialysis therapy between July 2003 and December 2006 from 54 centers in Korea



Am J Kidney Dis 54:711-720. © 2009 by the National Kidney Foundation, Inc.

Unmatched cohort



Matched cohort

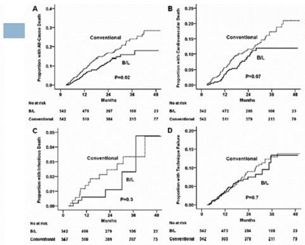


Table 3. Patient Outcomes in the Unmatched Cohort According to Different Dialysis Solutions

	B/L (n = 542)		Conventional (n = 1,621)			
Outcome	No. (%)	/1,000 Patient-years	No. (%)	/1,000 Patient-years	Hazard Ratio (95% confidence interval)	*P*
All-cause death	52 (9.6)	59.3	335 (20.7)	98.8	0.69 (0.52-0.93)	0.02
Cardiovascular death	36 (6.6)	41.0	225 (13.9)	66.4	0.71 (0.50-1.02)	0.06
Infection-related death	6 (1.1)	6.8	46 (2.8)	15.6	0.64 (0.27-1.52)	0.3
Technique failure	33 (6.1)	37.6	145 (8.9)	42.8	0.98 (0.67-1.44)	0.9

Abbreviations and definitions: B/L, physiological-pH bicarbonate/lactate-buffered peritoneal dialysis solution; conventional, lactate-buffered glucose-based peritoneal dialysis solution.

Table 5. Patient Outcomes in the Matched Cohort According to Different Dialysis Solutions

	B/L (n = 542)		Conventional (n = 542)			
Outcome	No. (%)	/1,000 Patient-years	No. (%)	/1,000 Patient-years	Hazard Ratio (95% confidence interval)	*P*
All-cause death	52 (9.6)	59.3	102 (18.9)	88.6	0.70 (0.50-0.98)	0.04
Cardiovascular death	36 (6.6)	41.0	70 (12.9)	60.8	0.73 (0.48-1.10)	0.1
Infection-related death	6 (1.1)	6.8	14 (2.6)	12.2	0.58 (0.22-1.52)	0.3
Technique failure	33 (6.1)	37.6	48 (8.9)	41.7	0.91 (0.58-1.43)	0.7

Abbreviations and definitions: B/L, physiological-pH bicarbonate/lactate-buffered peritoneal dialysis solution; conventional, lactate-buffered glucose-based peritoneal dialysis solution.

Table 4. Multivariate Cox Proportional Hazard Models for Mortality

	Unmatched Cohort		Matched Cohort	
	Hazard Ratio (95% confidence interval)	*P*	Hazard Ratio (95% confidence interval)	*P*
B/L (*v* conventional)	0.69 (0.52-0.93)	0.02	0.70 (0.50-0.98)	0.04
Sex (women *v* men)	1.16 (0.95-1.42)	0.2	1.05 (0.76-1.45)	0.8
Age (/1-y increase)	1.07 (1.06-1.08)	<0.001	1.07 (1.05-1.09)	<0.001
Diabetes (*v* no diabetes)	1.99 (1.58-2.49)	<0.001	1.91 (1.35-2.70)	<0.001
Cardiovascular comorbidity (*v* no)	1.36 (1.06-1.76)	0.02	1.31 (0.87-1.98)	0.2
Low socioeconomic status (*v* high)	1.16 (0.91-1.46)	0.2	1.35 (0.95-1.92)	0.1
Center (*v* less experienced)	0.83 (0.62-1.12)	0.2	0.49 (0.32-0.76)	0.001
Icodextrin (*v* no use)	0.43 (0.34-0.55)	<0.001	0.40 (0.28-0.58)	<0.001

Note: Models were adjusted for age, sex, diabetes, cardiovascular comorbidity, socioeconomic status, center experience, and icodextrin use.

Abbreviations and definitions: B/L, physiological-pH bicarbonate/lactate-buffered peritoneal dialysis solution; conventional, lactate-buffered glucose-based peritoneal dialysis solution.

Software Packages

- SAS: SUGI 214-26 "GREEDY" Macro
- STATA: PSMATCH2
  - (developed by Edwin Leuven and Barbara Sianesi [2003], as a user-supplied routine in STATA)
  - the most comprehensive package that allows users to fulfill most tasks for propensity score matching
- R: MatchIT
- SPSS: Matching을 하기 힘들다…

SAS® MACROS

%GREEDMTCH does nearest neighbor within caliper matching. If more than one control matches to a
case, the control is selected at random. No replacement is adopted in this macro.
http://www2.sas.com/proceedings/sugi26/p214-26.pdf

%OneToManyMTCH allows propensity score matching from 1-to-1 to 1-to-N based on specification from user. In use of nearest neighbor algorithm, the macro makes "best" matches first and "next-best" matches next in a hierarchical sequence until all matched controls use up. Each control can be selected once. This macro contains some ideas of greedy matching.
http://www2.sas.com/proceedings/sugi29/165-29.pdf

%Mahalanobis and **%MATCH** was proposed by Feng and his colleges at Eli Lilly Inc. in 2006. It implements the nearest available mahalanobis metric matching within calipers determined by propensity score. The caliper is defined as one quarter of standard deviation of the logit of propensity
score. %Mahalanobis is nested in %MATCH for mahalanobis distance calculation when there are more than two matched control subjects.
http://www.lexjansen.com/pharmasug/2006/publichealthresearch/pr05.pdf

SAS® MACROS

%MATCHUP was proposed by Martin and Ganguly, where both mahalanobis metric matching within caliper and a simple matching within caliper are provided with the caliper defined by the logit of propensity score. If processing time or capacity is a concern, a simple alternative can be performed quickly compared to the mahalanobis method at the potential price of more bias.
http://www.rx.uga.edu/main/home/cas/faculty/propensity.pdf

%BLINPLUS was developed using a validation study containing data on the two estimated propensity scores (error-prone and gold-standard) as well as the parameter estimates, their standard errors and covariances from Cox proportional hazards model. The error-prone propensity score was estimated from immeasurable confounding factors. The macro output provides the adjusted relative hazard rate estimates, including 95% confidence intervals adjusted for additional uncertainty from the estimation of the error-prone model (Stürmer and coworkers, 2005).
http://www.hsph.harvard.edu/faculty/spiegelman/blinplus/blinplus8.sas

SAS® MACROS

%MATCH was developed by Bergstralh and his coworkers in 1995 and replaced by %GMATCH, %VMATCH and %DIST later. Using distance measure, this macro can be used to match 1 or more controls for each subject in the treated group by means of greedy or optimal algorithms. %MATCH worked originally
on observational variables before the propensity score was introduced into the model as a matching factor.
http://mayoresearch.mayo.edu/mayo/research/biostat/sasmacros.cfm.

%MCSTRAT was developed by Vierkant and his colleges in 2000. This macro fits a conditional logistic model from matched or finely stratified data by propensity score and generates tables to describe the matching results as well as independent variables included in model. PROC PHREG and PROC IML are two major SAS procedures that perform regression diagnoses such as leverage value, delta chi-square, influential statistics and several other tools.
http://mayoresearch.mayo.edu/mayo/research/biostat/upload/65.pdf

S-PLUS/R PACKAGES

Matching by Sekhon provides functions for estimating causal effects by multivariate and propensity score matching and for finding best balance based on genetic search algorithm. The package includes a variety of univariate and multivariate tests to determine if balance has been obtained by the matching procedure.
http://jsekhon.fas.harvard.edu/matching

MATCHIT. This technique matches each treated unit to all possible control units with exactly the same value on all the covariates and propensity score by a variety of matching algorithms including exact matching, subclassification matching, nearest neighbor matching, optimal matching, full matching and genetic matching. After matching, balance is checked using plot or several other tools. The imputation of missing values is done via simulation using a parametric model together with Monte Carlo estimation. The subsequent analyses based on a parametric statistical model reduces the model dependence on background noises.
http://gking.harvard.edu/matchit/docs/matchit.html.

S-PLUS/R PACKAGES

OPTMATCH. This is bipartite matching, matching algorithms fall into two categories, greedy and optimal, where the latter functions with reconsideration of previous made matches in each new step. It combines propensity score with other multivariate distances, and functions for adaptations for large and memoryintensive issues. Full matching is its core algorithm, matching ratio between treated and control can vary.
http://stat.cmu.edu/R/CRAN/doc/packages/optmatch.pdf

USPS. This is one of packages developed by Ben Obenchain, containing R functions that perform a variety of alternative approaches to propensity scoring analyses. The USPS methods implement various forms of a-posteriori matching or stratification of patients who received only one of the two treatments that are being compared.
http://www.math.iupui.edu/~indyasa/.

S-PLUS/R PACKAGES

Propensity score profiling (Propscore) This software package is for comparing among multiple groups using propensity score with shrinkage estimation while addressing regression to the mean that can result from such multiple comparisons, based on Bayesian principles.
http://www.biostat.jhsph.edu/~cfrangak/papers/proscore_profiling/routine.r

Twang toolkit. Among the procedures available in the Twang toolkit, PS calculates propensity scores and diagnoses them using a variety of methods, centered on using boosted logistic regression. DIAG plot creates the diagnostic plots of propensity score; SENSITIVITY does sensitivity analysis, to name a few.
http://rss.acs.unt.edu/Rdoc/library/twang/html/00Index.html

Software Packages

- There is currently no commercial software package that offers formal procedure for PSM.
  - In SAS, Lori Parsons developed several Macros (e.g., the **GREEDY macro** does **nearest neighbor within caliper matching**).
  - In SPSS, Dr. John Painter of Jordan Institute developed a SPSS macro to do similar works as GREEDY (http://sswnt5.sowo.unc.edu/VRC/Lectures/index.htm).
- We have investigated several computing packages and found that PSMATCH2 (developed by Edwin Leuven and Barbara Sianesi [2003], as a user-supplied routine in STATA) is the most comprehensive package that allows users to fulfill most tasks for propensity score matching, and the routine is being continuously improved and updated.

Others

STATA: PSMATCH2, http://ideas.repec.org/c/boc/bocode/s432001.html.
PSCORE, ATTND, ATTNW, ATTR, ATTS AND ATTK, http://www.labor-torino.it/pdf_doc/ichino.pdf **MATCH**
http://elsa.berkeley.edu/~imbens/statamatching.pdf

SPSSPropensityMatching Macro.
http://sswnt5.sowo.unc.edu/VRC/Lectures/index.htm

S-plus with FORTRAN Routine

Limitations of Propensity Scores

- Large samples are required
- Group overlap must be substantial
- Hidden bias may remain because matching only controls for observed variables (to the extent that they are perfectly measured)

(Shadish, Cook, & Campbell, 2002)

Criteria for "Good" PSM

- Identify treatment and comparison groups with substantial overlap
- Match, as much as possible, on variables that are precisely measured and stable (to avoid extreme baseline scores that will regress toward the mean)
- Use a composite variable—e.g., a propensity score—which minimizes group differences across many scores