



**Center for Health Care Research & Policy**  
Case Western Reserve University / MetroHealth Medical Center  
Cleveland, Ohio



## **Reducing the Impact of Selection Bias with Propensity Scores**

**7<sup>th</sup> International Conference on Health Policy Statistics [ICHPS]  
Philadelphia, January 18, 2008**

Thomas E. Love, Ph.D. [Thomas.Love@case.edu](mailto:Thomas.Love@case.edu)

Center for Health Care Research and Policy  
Case Western Reserve University at MetroHealth Medical Center  
Cleveland, Ohio [www.chrp.org](http://www.chrp.org)

# Partially Annotated Bibliography for ICHPS 2008 Workshop

## REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

### References I Use in Teaching A Lot

1. Ahmed, A, Husain A, Love TE et al. (2006). "Heart failure, chronic diuretic use, and increase in mortality and hospitalization: An observational study using propensity scores." *European Heart Journal* 27: 1431-1439. [Uses propensity score matching combined with survival analyses to assess the impact of diuretics in HF patients. Incorporates formal sensitivity analysis and careful assessment of covariate balance, including the standardized differences plot.]
2. Austin, P. C. and M. M. Mamdani (2006). "A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use." *Statistics in Medicine* 25: 2084-2106. [Examines multiple approaches to propensity score modeling applied to the same data, including matching, stratification, covariate adjustment and weighting. **A great "second paper" to read.** Also looks at several ways of assessing residual confounding after stratification or matching.]
3. Connors, A. F., Jr., T. Speroff, et al. (1996). "The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators." *Jama* 276(11): 889-97. [Detailed PS matching and multivariate adjustment to account for selection bias, including sensitivity analysis - state of the art in 1996.]
4. D'Agostino, R. B., Jr. (1998). "Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group." *Statistics in Medicine* 17: 2265-2281. [Very readable introduction to matching, stratification, and regression using PS, with useful examples. **Excellent "first paper" to read.**]
5. Gum, P. A., M. Thamarasan, et al. (2001). "Aspirin Use and All-Cause Mortality Among Patients Being Evaluated for Known or Suspected Coronary Artery Disease: A Propensity Analysis." *JAMA* 286: 1187-1194. [PS Application: PS matching with non-proportional hazards model for survival analysis - results show an interesting impact of selection bias adjustment. Editorial by MJ Radford and JM Foody "How Do Observational Studies Expand the Evidence Base for Therapy?" pp. 1228-1230 - motivates observational studies vs RCT discussion, clustering, hierarchical models]
6. Joffe, M. M. and P. R. Rosenbaum (1999). "Invited commentary: propensity scores." *Am J Epidemiol* 150(4): 327-33. [Motivation for PS approaches geared to physician-epidemiologists, as well as a set of extensions - to case/control studies and to dose/response issues]
7. **Rosenbaum, P. R. (2002). *Observational Studies*. New York, Springer. [The major recent work on the subject of observational studies.]**
8. Rosenbaum, P. R. (1991). "Discussing hidden bias in observational studies." *Ann Intern Med* 115: 901-5. [Some of the fundamentals of sensitivity analysis]
9. Rosenbaum, P. R. and D. B. Rubin (1985). "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score." *The American Statistician* 39(1): 33-38. [Seminal paper on propensity matching]
10. Rubin, D. B. (1997). "Estimating Causal Effects from Large Data Sets Using Propensity Scores." *Ann Intern Med* 127 (Part 2): 757-763. [A very readable introductory discussion of key propensity score issues.]
11. Rubin, D. B. (2001). "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation." *Health Services & Outcomes Research Methodology* 2: 169-188. [Detailed (yet readable) description of PS matching, subclassification, weighting, under high stakes. **Excellent "third paper" to read.**]

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

#### Other References of Interest

12. Austin, P. C., M. M. Mamdani, et al. (2005). "The use of the propensity score for estimating treatment effects: Administrative versus clinical data." *Statistics in Medicine* 24: 1563-1578. [Stratifying on PS quintiles derived from administrative data did not also successfully balance clinical patient characteristics. Also, clinical data-based estimates of treatment effectiveness were attenuated compared to estimates from administrative data.]
13. Barker, F. G., II, S. Amin-Hanjani, et al. (2004). "Age-Dependent Differences in Short-Term Outcome after Surgical or Endovascular Treatment of Unruptured Intracranial Aneurysms in the United States, 1996-2000." *Neurosurgery* 54: 18-30. [PS Application: Stratified treatment effect analysis by propensity score quintile. PS for surgical treatment modeled for 1780 patients, using demographics, clinical characteristics, primary payer for care, type of admission, and hospital characteristics.]
14. Barnard, J., C. E. Frangakis, et al. (2003). "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City." *Journal of the American Statistical Association* 98(462): 299-323, including Comments and Rejoinder. [Application of the "principal stratification" framework to the problems of noncompliance and missing data in a randomized trial of school vouchers.]
15. Barsky, R., J. Bound, et al. (2002). "Accounting for the Black-White Wealth Gap: A Nonparametric Approach." *Journal of the American Statistical Association* 97(459): 663-673. [Interested in the effect of being black on "wealth-holding" in the black population. Approach involves reweighting the white wealth distribution based on a non-parametric estimate of the propensity score.]
16. Beddhu, S., M. H. Samore, et al. (2003). "Impact of timing of initiation of dialysis on mortality." *J Am Soc Nephrol* 14(9): 2305-12. [PS Application: 3 PS strata created and applied to a series of stratified Cox PH models. Some sensitivity analyses to account for the impact of censoring.]
17. Berg, G. D., A. Johnson, et al. (2003). "Clinical and Utilization Outcomes for a Pediatric and Adolescent Telephonic Asthma Care Support Program: A Propensity Score-Matched Cohort Study." *Dis Manage Health Outcomes* 11(11): 737-743. [PS Application: PS matching to find a suitable comparison cohort for 318 asthma care support program participants based on demographics, utilization, and procedures performed.]
18. Berger, V. W. (2005). "The reverse propensity score to detect selection bias and correct for baseline imbalances." *Statistics in Medicine* 24: 2777-2787. [Deals with the problem of potentially systematic baseline imbalance in individual blinded randomized trials. The reverse PS (defined within) permits detection of and correction for selection bias.]
19. Bingenheimer, J. B., R. T. Brennan, et al. (2005). "Firearm violence exposure and serious violent behavior." *Science* 308(5726): 1323-6. [PS Application: Examines longitudinal data on Chicago adolescents using a 12-way PS stratification (PS below 0.05 or above 0.75 then excluded) to look at relationship between exposure to firearm violence and subsequent perpetration of serious violence.]
20. Braitman, L. E. and P. R. Rosenbaum (2003). "Rare Outcomes, Common Treatments: Analytic Strategies Using Propensity Scores." *Ann Intern Med* 137(8): 693-696. [Propensity Scores: What They Are, Why They Work, and What They Can't Do. With subsequent Letter from Longstreth WT, Fahrenbruch CE and Koepsell TD "Propensity Scores for Rare Outcomes and Common Treatments" v. 139, pp. 152-153 which describes the use of PS to deal with imbalances observed in RCTs which need to be summarized with risk differences.]
21. Cannon, C. P., G. M. Gibson, et al. (2000). "Relationship of Symptom-Onset-to-Balloon Time and Door-to-Balloon Time With Mortality in Patients Undergoing Angioplasty for Acute Myocardial Infarction." *JAMA* 283: 2941-2947. [PS Application: PS used to address possible confounding - no real detail, but led to somewhat reduced risk increase with

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

- increases in door-to-balloon time. With Editorial by Michael S. Lauer "Primary Angioplasty - Time Is of the Essence" with several positive comments about the PS, also identifying need for lots of data and the fact that PS does nothing about hidden biases.]
22. Cepeda, M. S., R. Boston, et al. (2003). "Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders." *American Journal of Epidemiology* 158(3): 280-287. [Simulations comparing standard logistic regression for the outcome incorporating a group of covariates to standard logistic regression incorporating only the exposure variable and the quintiles of PS, based on the same covariates. Not that interesting a comparison, in my view. Conclusion was that # of events per confounder was the issue (small -> PS, >8 ->logistic regression).]
  23. Christakis, N. A. and T. J. Iwashyna (2003). "The health impact of health care on families: a matched cohort study of hospice use by decedents and mortality outcomes in surviving, widowed spouses." *Soc Sci Med* 57(3): 465-75. [PS Application: PS matching used to match couples where decedent was in hospice to couples where decedent was not.]
  24. Chu, J. H., S. J. Gange, et al. (2005). "Hormonal contraceptive use and the effectiveness of highly active antiretroviral therapy." *Am J Epidemiol* 161(9): 881-90. [PS Application: 1:1 PS match (actually they first stratified, found no mean differences within most strata in terms of PS, then selected matches at random from within the stratum, they used nearest-neighbor in the first stratum) of 77 hormonal contraceptive users to 77 non-users on the basis of demographics, clinical and lab characteristics, and prior use of health-care services.]
  25. Clark, C. and A. R. Rich (2003). "Outcomes of homeless adults with mental illness in a housing program and in case management only." *Psychiatr Serv* 54(1): 78-83. [PS Application: PS subclassification to form high-, medium- and low-impairment groups based on psychiatric symptoms, alcohol, and drug use.]
  26. Cohen, G. (2005). "Propensity-Score Methods and the Lenin School." *J of Interdisciplinary History* 36(2): 209-232. [PS Application: PS matching to assess whether study at the International Lenin School in Moscow increased the chances of British students assuming the leadership roles that the Communist International (Comintern) expected of them. Confronts in some detail the issue of estimating PS using non-ignorably missing data.]
  27. Cole, J. A., J. E. Loughlin, et al. (2002). "The effect of zanamivir treatment on influenza complications: a retrospective cohort study." *Clin Ther* 24(11): 1824-39. [PS Application: PS matching used to identify comparison groups with similar health service utilization, comorbidities and diagnosis (influenza) but no antiviral therapy.]
  28. Copas, A. J. and V. T. Farewell (1998). "Dealing with Non-Ignorable Non-Response by Using an "Enthusiasm-To-Respond" Variable." *Journal of the Royal Statistical Society, Series A* 161(3): 385-396. [Development of a "propensity-to-respond" score based on an embarrassment assessment combined with demographics. Applied to an estimation of virginity from the National Survey of Sexual Attitudes and Lifestyles.]
  29. Craske, M. G., D. Golinelli, et al. (2005). "Does the addition of cognitive behavioral therapy improve panic disorder treatment outcome relative to medication alone in the primary-care setting?" *Psychol Med* 35(11): 1645-54. [PS Application: PS Weighting, using a generalized boosted model built from confounders including study site, insurance, demographics, and various psychological and clinical measures. Only retained significant bivariate differences in PS model.]
  30. Crown, W. H., R. L. Obenchain, et al. (1998). "The application of sample selection models to outcomes research: The case of evaluating the effects of antidepressant therapy on resource utilization." *Statistics in Medicine* 17: 1943-1958. [Illustration of sample selection models, which attempt to control hidden biases in treatment selection which are also correlated with the outcome of interest.]
  31. D'Agostino, R., Jr. and D. B. Rubin (2000). "Estimating and using propensity scores with partially missing data." *Journal of the American Statistical Association* 95: 749-759.

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

32. D'Agostino, R. B., Jr., W. Lang, et al. (2001). "Examining the Impact of Missing Data on Propensity Score Estimation in Determining the Effectiveness of Self-Monitoring of Blood Glucose (SMBG)." *Health Services & Outcomes Research Methodology* 2: 291-315. [Assesses several PS-based and ANCOVA techniques for dealing with missing data in estimating the PS, and evaluating the outcome.]
33. Dehejia, R. H. (2004). "Practical propensity score matching: A reply to Smith and Todd." *J of Econometrics* 125: 355-364. [Discusses PS matching (1:1) in response to Smith and Todd's paper on matching for non-experimental data. Points out that a separate PS specification should be estimated for each comparison that you want to make, and that sensitivity analyses are very important, as well.]
34. Dehejia, R. H. and S. Wahba (2002). "Propensity Score-Matching Methods for Nonexperimental Causal Studies." *The Review of Economics and Statistics* 84(1): 151-161. [What to do when few controls are comparable to treated subjects. Uses data from the National Supported Work experiment and PS matching, comparing treatment effect estimates obtained from matching on the PS to benchmark results obtained from an experiment.]
35. Earle, C. C., J. S. Tsai, et al. (2001). "Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis." *J Clin Oncol* 19(4): 1064-70. [Interesting combination of instrumental variables and propensity score analyses.]
36. El-Bassel, N., L. Gilbert, et al. (2005). "Relationship between drug abuse and intimate partner violence: a longitudinal study among women receiving methadone." *Am J Public Health* 95(3): 465-70. [PS Application: Matching after multiple imputation using demographics, history of trauma, psychological distress, social support and HIV risks. Data were collected in three waves: confounders measured pre-treatment, outcomes post-treatment.]
37. Foody, J. M., C. R. Cole, et al. (2001). "A propensity analysis of cigarette smoking and mortality with consideration of the effects of alcohol." *Am J Cardiol* 87(6): 706-11. [PS Application: PS incorporated as a continuous variable in Cox PH models, after dropping largest and smallest quintiles.]
38. Foster, E. M. (2003). "Propensity Score Matching: An Illustrative Analysis of Dose Response." *Medical Care* 41(10): 1183-1192. [Use of PS methods for health services researchers, PS matching as a weighting for dealing with more than two exposure groups.]
39. Greenland, S. (2000). "An introduction to instrumental variables for epidemiologists." *International Journal of Epidemiology* 29: 722-729. [Introductory discussion of instrumental variables, for epidemiologists, with an application dealing with non-parametric adjustment for non-compliance in randomized trials.]
40. Greenland, S. and H. Morgenstern (2001). "Confounding in Health Research." *Annu. Rev. Public Health* 22: 189-212. [Describes confounding the context of a counterfactual model of causation - source of bias in the estimation of causal effects.]
41. Gu, X. S. and P. R. Rosenbaum (1993). "Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms." *J of Computational and Graphical Statistics* 2(4): 405-420. [Optimal matching beats greedy (nearest available) matching, sometimes by quite a lot, in getting closely matched pairs, but not by much in terms of balancing the matched samples. Optimal full matching is substantially better than Optimal 1:k matching.]
42. Hahn, J. (1998). "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects." *Econometrica* 66(2): 315-331.
43. Hansen, B. B. (2004). "Full Matching in an Observational Study of Coaching for the SAT." *Journal of the American Statistical Association* 99(467): 609-618.
44. Harraf, F., A. K. Sharma, et al. (2002). "A multicentre observational study of presentation and early assessment of acute stroke." *BMJ* 325: 17-21.
45. Heckman, J. J., N. Hohmann, et al. (2000). "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment." *The Quarterly Journal of Economics* CXV: 651-694.

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

46. Heckman, J. J., H. Ichimura, et al. (1997). "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64(4): 605-654.
47. Henry, A. D., A. M. Lucca, et al. (2004). "Inpatient hospitalizations and emergency service visits among participants in an Individual Placement and Support (IPS) model program." *Ment Health Serv Res* 6(4): 227-37.
48. Heuschmann, P. U., K. Berger, et al. (2003). "Frequency of thrombolytic therapy in patients with acute ischemic stroke and the risk of in-hospital mortality: the German Stroke Registers Study Group." *Stroke* 34(5): 1106-13. [PS Application: Patients matched by propensity for tPA therapy in the context of comparing risk of inpatient death.]
49. Hill, A. B., D. Obrand, et al. (2000). "Hemispheric Stroke following Cardiac Surgery: A Case-Control Estimate of the Risk Resulting from Ipsilateral Asymptomatic Carotid Artery Stenosis." *Ann Vasc Surg* 14: 200-209. [PS Application: Case-control study with PS stratification - discussion includes a meta-analysis of related prospective cohort studies.]
50. Hirano, K. and G. W. Imbens (2001). "Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services & Outcomes Research Methodology* 2: 259-278.
51. Hirano, K., G. W. Imbens, et al. (2003). "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71(4): 1161-1189.
52. Hodges, K. and H. Grunwald (2005). "The use of propensity scores to evaluate outcomes for community clinics: identification of an exceptional home-based program." *J Behav Health Serv Res* 32(3): 294-305.
53. Holden, C. (2005). "Sociology. Controversial study suggests seeing gun violence promotes it." *Science* 308(5726): 1239-40. [Comment on Bingenheimer et al.]
54. Huang, I.-C., C. Frangakis, et al. (2005). "Application of a Propensity Score Approach for Risk Adjustment in Profiling Multiple Physician Groups on Asthma Care." *Health Services Research* 40(1): 253-278.
55. Hullsiek, K. H. and T. A. Louis (2002). "Propensity score modeling strategies for the causal analysis of observational data." *Biostatistics* 2(4): 179-193.
56. Huncharek, M., B. Kupelnick, et al. (2002). "Maternal smoking during pregnancy and the risk of childhood brain tumors: a meta-analysis of 6566 subjects from twelve epidemiological studies." *J Neurooncol* 57(1): 51-7.
57. Ichimura, H. and C. Taber (2001). "Propensity-Score Matching with Instrumental Variables." *The American Economic Review* 91(2): 119-124.
58. Imai, K. and D. A. van Dyk (2004). "Causal Inference with General Treatment Regimes: Generalizing the Propensity Score." *Journal of the American Statistical Association* 99(467): 854-866.
59. Imbens, G. W. (2000). "The role of the propensity score in estimating dose-response functions." *Biometrika* 87(3): 706-710.
60. Jaar, B. G., J. Coresh, et al. (2005). "Comparing the Risk for Death with Peritoneal Dialysis and Hemodialysis in a National Cohort of Patients with Chronic Kidney Disease." *Ann Intern Med* 143: 174-183. [Comment by Gerald Schuman, "Mortality and Treatment Modality of End-Stage Renal Disease"]
61. Joffe, M. M., T. R. Ten Have, et al. (2004). "Model Selection, Confounder Control, and Marginal Structural Models: Review and New Applications." *The American Statistician* 58(4): 272-279.
62. Jones, A. S., R. B. D'Agostino, Jr., et al. (2004). "Assessing the Effect of Batterer Program Completion on Reassault Using Propensity Scores." *J of Interpersonal Violence* 19(9): 1002-1020.

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

63. Katzan, I. L., R. D. Cebul, et al. (2003). "The effect of pneumonia on mortality among patients hospitalized for acute stroke." *Neurology* 60(4): 620-5. [PS Application: PS for pneumonia used for risk adjustment in logistic regression analyses.]
64. Kazmi, W. H., G. T. Obrador, et al. (2004). "Late nephrology referral and mortality among patients with end-stage renal disease: a propensity score analysis." *Nephrol Dial Transplant* 19(7): 1808-14.
65. Kennedy, J., H. Quan, et al. (2005). "Statins Are Associated With Better Outcomes After Endarterectomy in Symptomatic Patients." *Stroke* 36: 2072-2076.
66. Kilian, R. and M. C. Angermeyer (2005). "The effects of antipsychotic treatment on quality of life of schizophrenic patients under naturalistic treatment conditions: an application of random effect regression models and propensity scores in an observational prospective trial." *Qual Life Res* 14(5): 1275-89.
67. Kleiman, N. S., A. M. Lincoff, et al. (2000). "Early percutaneous coronary intervention, platelet inhibition with eptifibatide, and clinical outcomes in patients with acute coronary syndromes. PURSUIT Investigators." *Circulation* 101(7): 751-7.
68. Klungel, O. H., E. P. Martens, et al. (2004). "Methods to assess intended effects of drug treatment in observational studies are reviewed." *J Clin Epidemiol* 57(12): 1223-31.
69. Koch, C. G., F. Khandwala, et al. (2003). "Gender and outcomes after coronary artery bypass grafting: a propensity-matched comparison." *J Thorac Cardiovasc Surg* 126(6): 2032-43.
70. Kraemer, H. C., E. Stice, et al. (2001). "How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping, and Proxy Risk Factors." *Am J Psychiatry* 158: 848-856.
71. Kubal, C., A. K. Srinivasan, et al. (2005). "Effect of risk-adjusted diabetes on mortality and morbidity after coronary artery bypass surgery." *Ann Thorac Surg* 79(5): 1570-6.
72. Landrum, M. B. and J. Z. Ayanian (2001). "Causal Effect of Ambulatory Specialty Care on Mortality Following Myocardial Infarction: A Comparison of Propensity Score and Instrumental Variable Analyses." *Health Services & Outcomes Research Methodology* 2: 221-245. [Direct comparison of propensity score and instrumental variable approaches.]
73. Landrum, M. B., B. J. McNeil, et al. (1999). "Understanding variability in physician ratings of the appropriateness of coronary angiography after acute myocardial infarction." *J Clin Epidemiol* 52(4): 309-19.
74. Lash, T. L., R. A. Silliman, et al. (2000). "The effect of less than definitive care on breast carcinoma recurrence and mortality." *Cancer* 89(8): 1739-47.
75. Lauby, J. L., P. J. Smith, et al. (2000). "A community-level HIV prevention intervention for inner-city women: results of the women and infants demonstration projects." *Am J Public Health* 90(2): 216-22. [PS Application: PS subclassification and regression runs/residuals for change score analyses.]
76. Leon, A. C. and D. Hedeker (2005). "A mixed-effects quintile-stratified propensity adjustment for effectiveness analyses of ordered categorical doses." *Statistics in Medicine* 24: 647-658.
77. Lieberman, E., J. M. Lang, et al. (1996). "Association of epidural analgesia with cesarean delivery in nulliparas." *Obstet Gynecol* 88(6): 993-1000.
78. Linden, A., J. L. Adams, et al. (2005). "Using Propensity Scores to Construct Comparable Control Groups for Disease Management Program Evaluation." *Dis Manage Health Outcomes* 13(2): 107-115.
79. Lindenauer, P. K., P. Pekow, et al. (2005). "Perioperative Beta-Blocker Therapy and Mortality after Major Noncardiac Surgery." *N Engl J Med* 353: 349-361. [Comment by Don Poldermans and Eric Boersma, "Beta-Blocker Therapy in Noncardiac Surgery"]
80. Logar, C. M., L. M. Pappas, et al. (2005). "Surgical revascularization versus amputation for peripheral vascular disease in dialysis patients: a cohort study." *BMC Nephrology* 6: 3.

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

81. Loughlin, J. E., J. A. Cole, et al. (2002). "Comparison of Resource Utilization by Patients Treated with Transdermal Fentanyl and Long-Acting Oral Opioids for Nonmalignant Pain." *Pain Medicine* 3(1): 47-55. [PS matching, tests for differences in costs, visits, etc.]
82. Love, T. E., R. D. Cebul, et al. (2003). "Effect of matching for propensity on the balance of 'unmeasured' covariates (Abstract)." *J Clin Epidemiol* 56(9): 920.
83. Lu, B. (2005). "Propensity Score Matching with Time-Dependent Covariates." *Biometrics* 61: 721-28.
84. Lu, B., E. Zanutto, et al. (2002). "Matching with doses in an observational study of a media campaign against drug abuse." *Journal of the American Statistical Association* 96: 1245-1253.
85. Lunceford, J. K. and M. Davidian (2004). "Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study." *Statistics in Medicine* 23: 2937-2960.
86. McCaffrey, D. F., G. Ridgeway, et al. (2004). "Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies." *Psychological Methods* 9(4): 403-425.
87. Mitra, N. and A. Indurkha (2005). "A propensity score approach to estimating the cost-effectiveness of medical therapies from observational data." *Health Economics* 14: 805-815.
88. Mitra, N., F. R. Schnabel, et al. (2001). "Estimating the Effect of an Intensive Surveillance Program on Stage of Breast Carcinoma at Diagnosis: A Propensity Score Analysis." *Cancer* 91: 1709-1715.
89. Morant, S. V., D. Pettitt, et al. (2004). "Application of a propensity score to adjust for channelling bias with NSAIDs." *Pharmacoepidemiology and Drug Safety* 13: 345-353.
90. Mortensen, E. M., M. I. Restrepo, et al. (2005). "The effect of prior statin use on 30-day mortality for patients hospitalized with community-acquired pneumonia." *Respiratory Research* 6: 82. ["A propensity score technique was used to balance covariates associated with statin use between groups {ref to Stone et al (1995) *Med Care*.} The covariates used were the pneumonia severity index score (which includes comorbid conditions such as congestive heart failure, liver disease, and history of stroke), history of alcoholism, history of diabetes mellitus, coronary artery disease, and current tobacco use. Variables were entered, and maintained, in the model if they had a p-value <0.20 in the univariate analysis (with statin use as the dependent variable) and had a p-value <0.20 in the final model."]
91. Moustaki, I. and M. Knott (2000). "Weighting for item non-response in attitude scales by using latent variable models with covariates." *Journal of the Royal Statistical Society, Series A* 163(3): 445-459.
92. Murray, P. K., T. E. Love, et al. (2005). "Rehabilitation services after the implementation of the nursing home prospective payment system: differences related to patient and nursing home characteristics." *Med Care* 43(11): 1109-15.
93. Murray, P. K., M. Singer, et al. (2003). "Outcomes of rehabilitation services for nursing home residents." *Arch Phys Med Rehabil* 84: 1129-1136. [PS Application: PS for rehab using 112 variables, with matching and multivariate adjustment (including dose-response) with detailed evaluation of effectiveness of the PS model and sensitivity analyses.]
94. Nash, I. S., R. R. Corrato, et al. (1999). "Generalist Versus Specialist Care for Acute Myocardial Infarction." *Am J Cardiol* 83: 650-654.
95. Normand, S.-L. T., M. B. Landrum, et al. (2001). "Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores." *J Clin Epidemiol* 54(4): 387-98. [PS Application: PS Matching with calipers - interesting discussion of covariate balance (standardized differences) - outcome assessment through odds ratios, Mantel-Haenszel and survival rates.]
96. O'Connor, C. M., E. J. Velazquez, et al. (2002). "Comparison of Coronary Artery Bypass Grafting Versus Medical Therapy on Long-Term Outcome in Patients With Ischemic



## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

- Cardiomyopathy (A 25-Year Experience from the Duke Cardiovascular Disease Databank)." *Am J Cardiol* 90: 101-107.
97. O'Keefe, S. (2004). "Job creation in California's enterprise zones: A comparison using a propensity score matching model." *J of Urban Economics* 55: 131-150.
  98. Petersen, L. A., S. L. Normand, et al. (2000). "Outcome of myocardial infarction in Veterans Health Administration patients as compared with medicare patients." *N Engl J Med* 343(26): 1934-41. [PS Application: Presents both PS matching and the standard multivariate adjustments (through logistic regression).]
  99. Peterson, J. G., E. J. Topol, et al. (2001). "Prognostic importance of concomitant heparin with eptifibatide in acute coronary syndromes. PURSUIT Investigators. Platelet Glycoprotein IIb/IIIa in Unstable Angina: Receptor Suppression Using Integrilin Therapy." *Am J Cardiol* 87(5): 532-6. [Dropped top two PS quintiles in this secondary analysis of an RCT - then included PS in Cox PH models.]
  100. Posner, M. A., A. S. Ash, et al. (2001). "Comparing Standard Regression, Propensity Score Matching, and Instrumental Variables Methods for Determining the Influence of Mammography on Stage of Diagnosis." *Health Services & Outcomes Research Methodology* 2: 279-290. [Comparison of PS matching, instrumental variables, and standard risk adjustment techniques for dealing with selection bias.]
  101. Potosky, A. L., J. Legler, et al. (2000). "Health outcomes after prostatectomy or radiotherapy for prostate cancer: results from the Prostate Cancer Outcomes Study." *J Natl Cancer Inst* 92(19): 1582-92. [PS Application: Direct PS adjustment in cross-sectional and longitudinal regression analyses - conclusions section motivates a study of hidden biases.]
  102. Reiter, J. (2000). "Using Statistics to Determine Causal Relationships." *The American Mathematical Monthly* 107(1): 24-32.
  103. Rosenbaum, P. R. (1987). "Model-Based Direct Adjustment." *Journal of the American Statistical Association* 82(398): 387-394.
  104. Rosenbaum, P. R. (1991). "Sensitivity Analysis for Matched Case-Control Studies." *Biometrics* 47(1): 87-100.
  105. Rosenbaum, P. R. (1999). "Choice as an Alternative to Control in Observational Studies." *Statistical Science* 14(3): 259-278.
  106. Rosenbaum, P. R. (1999). "Reduced Sensitivity to Hidden Bias at Upper Quantiles in Observational Studies with Dilated Treatment Effects." *Biometrics* 55(2): 560-564.
  107. Rosenbaum, P. R. (2001). "Effects attributable to treatment: Inference in experiments and observational studies with a discrete pivot." *Biometrika* 88(1): 219-231.
  108. Rosenbaum, P. R. (2005). "Heterogeneity and Causality: Unit Heterogeneity and Design Sensitivity in Observational Studies." *The American Statistician* 59(2): 147-152.
  109. Rosenbaum, P. R. and D. B. Rubin (1983). "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome." *Journal of the Royal Statistical Society, Series B* 45: 212-218.
  110. Rosenbaum, P. R. and D. B. Rubin (1984). "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79(387): 516-524. [Seminal paper - propensity score subclassification in detail.]
  111. Rosenbaum, P. R. and D. B. Rubin (1985). "The Bias Due to Incomplete Matching." *Biometrics* 41(1): 103-116. [A seminal paper - some thoughts on sensitivity analysis, and a detailed argument about why statisticians want to match everyone.]
  112. Rubin, D. B. (2004). "On principles for modeling propensity scores in medical research." *Pharmacoepidemiol Drug Saf* 13(12): 855-7.
  113. Rubin, D. B. and N. Thomas (1992). "Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Distributions." *Biometrika* 79(4): 797-809.
  114. Rubin, D. B. and N. Thomas (1996). "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics* 52(1): 249-264.

## Partially Annotated Bibliography for ICHPS 2008 Workshop

### REDUCING THE IMPACT OF SELECTION BIAS WITH PROPENSITY SCORES

115. Rubin, D. B. and N. Thomas (2000). "Combining propensity score matching with additional adjustments for prognostic covariates." *Journal of the American Statistical Association* 95: 573-585.
116. Seeger, J. D., P. L. Williams, et al. (2005). "An application of propensity score matching using claims data." *Pharmacoepidemiology and Drug Safety* 14: 465-476.
117. Shah, B. R., A. Laupacis, et al. (2005). "Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review." *J Clin Epidemiol* 58(6): 550-9.
118. Shlipak, M. G., W. S. Browner, et al. (2001). "Comparison of the effects of angiotensin converting-enzyme inhibitors and beta blockers on survival in elderly patients with reduced left ventricular function after myocardial infarction." *Am J Med* 110(6): 425-33. [PS Application: PS models for patients on beta-blockers, ACE inhibitors, or both, vs. neither (control). Classified by quintiles, yielding multiple regressions. With Editorial by MS Lauer "Medical Therapy for Coronary Artery Disease Works, Even (Especially) in the Real World"]
119. Silber, J. H., P. R. Rosenbaum, et al. (1995). "Comparing the Contributions of Groups of Predictors: Which Outcomes Vary with Hospital Rather Than Patient Characteristics." *Journal of the American Statistical Association* 90(429): 7-18.
120. Silber, J. H., P. R. Rosenbaum, et al. (2001). "Multivariate matching and bias reduction in the surgical outcomes study." *Med Care* 39(10): 1048-64. [Development of multivariate PS-matched case-control pairs using Medicare claims data.]
121. Smith, H. L. (1997). "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies." *Sociological Methodology* 27: 325-353.
122. Smith, J. A. and P. E. Todd (2005). "Does matching overcome LaLonde's critique of nonexperimental estimators?" *J of Econometrics* 125: 305-353.
123. Sturmer, T., S. Schneeweiss, et al. (2005). "Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration." *Am J Epidemiol* 162(3): 279-89.
124. Sturmer, T., S. Schneeweiss, et al. (2005). "Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly." *Am J Epidemiol* 161(9): 891-8.
125. Ten Have, T. R., M. R. Elliott, et al. (2004). "Encouragement Trials in Treating Primary Care Depression." *Journal of the American Statistical Association* 99(465): 16-25.
126. Tu, W. and X.-H. Zhou (2003). "A Bootstrap Confidence Interval Procedure for the Treatment Effect Using Propensity Score Subclassification." *Health Services & Outcomes Research Methodology* 3: 135-147.
127. Wang, J., P. T. Donnan, et al. (2001). "The multiple propensity score for analysis of dose-response relationships in drug safety studies." *Pharmacoepidemiology and Drug Safety* 10: 105-111.
128. Weitzen, S., K. L. Lapane, et al. (2004). "Principles for modeling propensity scores in medical research: a systematic literature review." *Pharmacoepidemiol Drug Saf* 13(12): 841-53.
129. Weitzen, S., K. L. Lapane, et al. (2005). "Weaknesses of goodness-of-fit tests for evaluating propensity score models: the case of the omitted confounder." *Pharmacoepidemiol Drug Saf* 14(4): 227-38.
130. Winkelmayr, W. C. and T. Kurth (2004). "Propensity scores: help or hype?" *Nephrol Dial Transplant* 19: 1671-1673.
131. Zacharias, A., T. A. Schwann, et al. (2005). "Is hospital procedure volume a reliable marker of quality for coronary artery bypass surgery? A comparison of risk and propensity adjusted operative and midterm outcomes." *Ann Thorac Surg* 79(6): 1961-9.