Using Propensity Score Matching in Clinical Investigations: A Discussion and Illustration

Carrie Hosman[1,*] and Hitinder S. Gurm[2]

[1]*From the Department of Statistics, University of Michigan, Ann Arbor MI, USA*

[2]*Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical Center, Ann Arbor, MI, USA*

Abstract: Propensity score matching is a useful tool to analyze observational data in clinical investigations, but it is often executed in an overly simplistic manner, failing to use the data in the best possible way. This review discusses current best practices in propensity score matching, outlining the method's essential steps, including appropriate post-matching balance assessments and sensitivity analyses. These steps are summarized as eight key traits of a propensity matched study. Further, this review illustrates these traits through a case study examining the impact of access site in percutaneous coronary intervention (PCI) procedures on bleeding complications. Through propensity score matching, we find that bleeding occurs significantly less often with radial access procedures, though many other outcomes show no significant difference by access site, a finding that mirrors the results of randomized controlled trials. Lack of attention to methodological principles can result in results that are not biologically plausible.

Keywords: Propensity Score Matching, Observational Data, Clinical Investigations.

For many clinical questions of interest, randomized controlled trials are impossible or their strict inclusion criteria limit generalization of RCT data to wider clinical practice. In such settings, when properly analyzed, good observational data can provide clinically useful information. A useful method for structuring and analyzing a study of observational data is propensity score matching. When done correctly, propensity score matching reduces the bias from residual confounding due to the lack of randomization to a certain exposure [1]. Unfortunately, many applications of propensity score matching in the medical literature use first generation techniques and skip key steps such as balance testing and sensitivity analysis [2]. Incorrect application of these methods has produced biologically implausible findings that are often at odds with results of randomized trials, leading to confusion and criticism of the technique. This article outlines proper use of propensity score matching for analyzing observational data by highlighting eight necessary criteria that should be met in creating the scores, matching, and performing subsequent analysis. A case study illustrates these points and the consequences of poor propensity score matching on an examination of the effects of access site location in percutaneous coronary intervention on post-procedure outcome variables. Meticulous attention to these basic principles can help prevent erroneous results and enhance the reliability of the findings.

*Address correspondence to this author at the University of Michigan, Department of Statistics, 439 West Hall, 1085 South University, Ann Arbor, MI 48109, USA; Tel: (248) 890-5621; Fax: (734) 764-4142;
E-mail: chosman@umich.edu

PROPENSITY SCORES

A propensity score reduces high-dimensional data to a single score, allowing an analyst to control for many more variables than could be included in a standard, regression-based, approach [1]. It is the probability a patient receives the exposure or treatment, conditional on observed confounders. Two patients with the same propensity score are, in theory, equally likely to receive the treatment condition. In a randomized experiment, the experimental protocol and randomization allow the propensity score to be known for each observation, but with observational data, the model and scores must be estimated. For binary exposure variables, a logistic regression model estimates the propensity score model. A propensity score is a balancing score, that is, treated and control subjects with the same propensity score have the same distribution of observed covariates [1]. In practice, a check of the validity of the balancing property provides the key diagnostic for the propensity score model and subsequent matching.

The first step in any propensity matching study should be a careful selection of the study population. Patients who, for clinical reasons will not be eligible for receiving either of the treatment conditions should be excluded. While this may seem obvious, this step can be frequently omitted and can result in major confounding that cannot be adjusted for by statistical methods.

A well-designed investigation using propensity scores only includes confounders in the propensity

E-ISSN: 1929-6029/15

score model that are covariates measured prior to the exposure or treatment; violating this principle challenges the conditional independence between the treatment assignment and outcome that defines a propensity score. A propensity score estimated with post-treatment variables may unintentionally describe part of the treatment effect in addition to the treatment assignment. While the goal of a propensity score model may appear to be to predict treatment assignment, the ultimate objective is to reduce the bias and increase the precision in estimated treatment effects. Therefore, an analyst need not only include variables that impact treatment assignment; covariates related to the outcome but not necessarily the treatment assignment should also be included [3,4]. In the case of missing data, mean imputation and the creation of indicator variables that describe the pattern of missing data should be used. This procedure does not change the fitted values of the propensity scores, though the coefficient values of the propensity score model will be affected [5]. When assessing a propensity score model, standard significance tests of the coefficients and the model are not useful. Further, measures like the c-statistic should not be used to assess a propensity score model: recent research has indicated there is no relationship between the c-statistic and the ability to balance prognostically important variables between treated and untreated subjects after matching [6]. In fact, the c-statistic can be increased to a point where few units would able to be matched. Rather than using standard regression diagnostics or the c-statistic, a good way to assess a propensity score model is to examine the resulting balance after matching is performed. If balance – the similarity in covariate distributions across treatment and control groups conditional on the propensity score or matched set – is poor, the propensity score model or matching should be adjusted, regardless of other model diagnostics.

Many registries have multi-centric data with variations in procedural preferences, and a propensity score matching should account for these differences. This adjustment can be accomplished by adding either a fixed or random effect for the center to the propensity score, a method called partial pooling by some authors [7]. Recent research [8] indicated that including a fixed effect in the propensity score model can achieve good balance in cluster-level observed variables after matching, and thus, it should theoretically achieve good balance in unobserved cluster-level variables. Additionally, propensity score models with either a fixed or random effect can both capture the unobserved heterogeneity that would exist if cluster-level effects are omitted. With multi-centric data, multilevel or fixed effect propensity score models perform better in terms of bias reduction in treatment effect estimates than a model that fails to account for clinic. After modification of the propensity score model, matching and subsequent analysis can proceed in the same manner as a study in which clinic-level features did not need to be considered.

After propensity scores are estimated, they could be used, without outcome information, to “design” an observational study [9]. Stratification, matching, and inverse probability of treatment weighting (IPTW) are all methods to use propensity scores without outcome information. Stratification is a coarsened matching that is most useful with a limited amount of data. Some argue that propensity score matching fails to utilize a large proportion of the data for which good matches cannot be found. This limitation can be overcome by the use of flexible matching methods like full matching [10,11], which will be discussed in the next section. A drawback of IPTW is that the propensity score model needs to be correctly specified as estimated propensity scores are used to weight cases. Covariate adjustment, in which the propensity scores are used in a regression model as a covariate, not only requires correctly specified propensity scores but also incorporates outcome information into the design stage of the observational study. For these reasons, this review will focus on matching as the preferred method for using propensity scores to “design” an observational study.

MATCHING ON THE PROPENSITY SCORE

There are many choices to make when matching: observations can be matched in an optimal or greedy (nearest available) manner; the number of treatment and control observations in each matched set can be variable or fixed (e.g., in the case of a pair match, these are fixed at one treatment and one control observation); or, restrictions can be placed on how different observations in the same matched set are permitted to be. (These restrictions, called calipers, are used to ensure greater similarity of units in a matched set.) Decisions to modify the matching method or propensity score model are best made following the creation of matched sets, and balance tests prove useful for this purpose.

Nearest available, or greedy, matching has been standard in propensity score matched studies, particularly in medical research. An observation is

matched to an observation with similar covariates, as measured by similarity in the propensity score, another pair is made similarly, and so on. This often results in the early matches being very good, but matches made toward the end of the matching procedure being poor, as the procedure must draw from an increasingly limited pool of leftover matches. Further, pair matching in general leaves a large reservoir of control observations unmatched and unused in subsequent analysis; ideally, a matching procedure should use these observations.

In contrast, optimal matching typically uses more data and creates a better arrangement of matched sets by considering all possible matching arrangements and choosing the arrangement that minimizes the cumulative disparity between observations across all matched sets [10,11]. If pair matching is desired, optimal pair matching is more favorable than nearest available pair matching. If a researcher can move beyond pair matching, optimal full matching allows observations to be placed into matched sets of varying sizes with differing numbers of treatment and control observations in each matched set. By matching in this flexible manner, the reservoir of control observations can be used by matching multiple control observations to a single treatment observation, discarding few observations. In addition, optimal full matching can lead to a greater reduction in bias than greedy matching [10,12]. Although more computing power may be required to make optimal matches than would be required to make nearest available pair matches with the same units, modern computers can handle the necessary algorithms efficiently.

BALANCE TESTS

In order to evaluate the propensity score model and the matching structure, the degree of disparity in covariate values across exposure and control groups should be assessed after matching. If a high degree of disparity exists after matching, then the propensity score model or matching structure should be adjusted. It is crucial that evaluations of balance incorporate the matching structure rather than simply assessing balance in the matched sample. If covariates are imbalanced between exposure and control groups, then any treatment effect discovered in subsequent analyses could simply be a result of differences in background variables.

While many authors have debated the merits of balance tests that rely on significance tests [13-15] we think they provide a straightforward assessment of the degree of disparity that remains after matching. One such method of balance tests outlined by Hansen and Bowers [15] and available in the R library `RItools` provides an omnibus chi-square test statistic to evaluate balance across all covariates with one test in addition to information about the imbalance in each covariate. This balance test uses randomization-based inference to assess balance on quantitative and categorical variables in a similar manner. A significant imbalance across all covariates indicates that the matching method or propensity score model needs modification.

SUBSEQUENT ANALYSIS USING PROPENSITY MATCHED SETS

Whether a model-based or nonparametric approach is taken for subsequent analysis, the subsequent analysis should account for the matched nature of the sample. Simply analyzing the matched sample is not sufficient as matching aims to replicate a block-randomized study in which treatment is assigned at random within matched sets. Ignoring this structure in a randomized experiment would subject a study to criticism; the same should be true in propensity matched observational studies. In a survey of propensity matched studies in medical literature [2], about 60% of the studies accounted for the matched nature of the data; of these studies that accounted for the matched structure, only about half of them handled the matched structure with an appropriate analytic method for the estimation of treatment effects. For binary outcomes, appropriate methods include weighted averages of matched differences, nonparametric tests such as Mantel-Haenszel tests, or conditional logistic regression in which the matched sets are the strata within which the coefficients are computed before aggregation.

SENSITIVITY ANALYSIS

After an analysis is conducted, forms of sensitivity analysis should be performed. First and foremost, biological plausibility should always be considered for any results obtained through analysis. Finding treatment to have an effect on an outcome unrelated to it indicates the potential presence of residual confounding that exists despite the statistical adjustment used. To check for potential residual confounding, examine outcomes known to be unrelated to a certain treatment in addition to the outcomes of specific interest. More formally, a sensitivity analysis

should be conducted to assess the sensitivity of the results to bias from unobserved variables [16]. These methods assess the robustness of research findings to residual confounding; findings that are significant but very sensitive to residual confounding may not be meaningful. In a survey of propensity matched studies in the medical literature [2], only one of 27 studies reported the results of a sensitivity analysis.

The method of sensitivity analysis used should depend on the analysis method. If subsequent analyses were nonparametric tests on matched data then, following Rosenbaum's method [16], one could conduct a sensitivity analysis by examining the effects of variations in a sensitivity parameter. For a range of values of this sensitivity parameter, which describes the relationship between the unmeasured confounder and treatment assignment, the sensitivity analysis provides bounds on the significance level for the original inference. On the other hand, if subsequent analyses utilized regression models to incorporate additional covariate information, methods that consider the model structure should be used [16-18]. These methods of sensitivity analysis consider specific values related to an unmeasured confounder, such as prevalence in the treated and control groups and a potential effect size, and computes confidence bounds for the resulting treatment effect. If the conclusions of the study are only altered for extreme values of the sensitivity parameter or hypothesized prevalence and effect size, then the study results are not sensitive to bias due to unobserved variables. A researcher can decide whether an unmeasured confounder could exist that would reverse any statistically significant findings from the previous analysis.

In the absence of a sensitivity analysis, the study can be questioned for the potential of the findings to be subject to confounding due to unmeasured variables. While this step of a propensity score matched study is often skipped, it is an important check of the robustness of a study's findings.

CASE STUDY

Through a case study, we illustrate the steps to conduct a propensity matched analysis considering the 8 key characteristics outlined in this comment. Throughout the case study, results of the preferred method are compared to some alternatives that may commonly be chosen in current medical research studies with propensity score matching.

Using a large database of information for percutaneous coronary interventions (PCIs) at participating hospitals in the state of Michigan, we examine the difference in post-procedure bleeding for femoral versus radial access for the PCI procedures. It is widely accepted that procedures using radial access should show a statistically significant impact on bleeding events, but the access locations should not present statistically significant differences in other outcomes, such as post-procedure heart failure, need for dialysis, or myocardial infarction. All analysis is performed using R version 3.0.2.

Data

This study includes data from patients undergoing PCI at 47 hospitals participating in the Blue Cross Blue Shield of Michigan Cardiovascular Consortium (BMC2). The details of the BMC2 registry and its data collection and auditing process have been described previously [19-21]. Procedural data on all patients undergoing PCI at participating hospitals are collected using standardized data collection forms. Baseline data include clinical, demographic, procedural, and angiographic characteristics as well as medications used before, during, and after the procedure, and in-hospital outcomes. All data elements have been prospectively defined, and the protocol has been either approved or the need for approval waived by local institutional review boards at each hospital.

The study population for this analysis included all consecutive patients who underwent PCI between October 1, 2012 and September 30, 2013 (n=30463), shown in Figure **1**. To compare femoral and radial access in terms of bleeding and other outcomes, some patients were excluded from analysis. These include patients who died in the catheterization laboratory (n=70), had a guiding catheter used that was too large for radial access (so the decision to use radial or femoral was already made, n=9461), and patients for whom the access site was not either not reported or was neither radial nor femoral (n=71). After exclusions, 20926 patients were included, of which 4854 (23.2%) had radial access PCI and 16072 (76.8%) had femoral access PCI.

Propensity Score and Matching

To estimate the propensity scores, we used a logistic regression model in which the covariates were demographics such as race, sex, and age, body mass index, PCI indication, and variables documenting

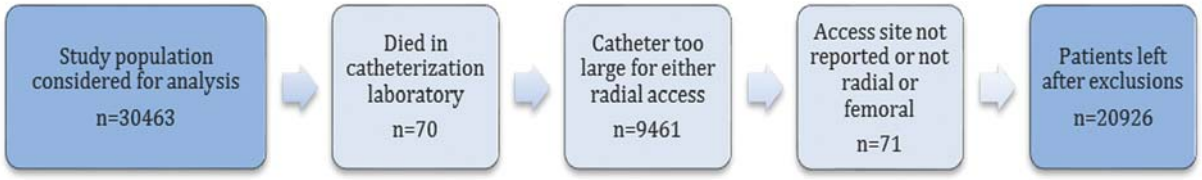

Figure 1: Study patient population and exclusions.

clinical comorbidities and medications prior to the PCI procedure. As the data come from 47 different hospitals, a fixed effect for hospital was included in the propensity score model. A comparison of the fitted propensity scores for the femoral and radial access groups is shown in Figure **2**. The support of the propensity scores is similar across the two groups, and the observations with propensity scores outside the range are left unmatched after matching.

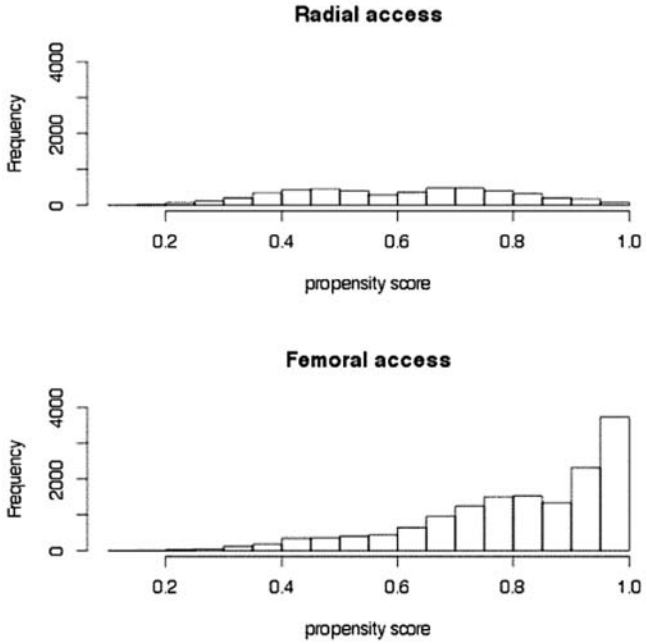

Figure 2: Comparison of propensity score distribution by access site.

After obtaining fitted propensity scores, we use the optmatch package in R to perform an optimal full match of the patients in the femoral access group to those in the radial access group. Two calipers - measuring 0.005 and 0.01 of one standard deviation in the propensity scores, respectively - were examined. For comparison, we also performed a greedy pair match to examine the balance and results from this match. As Table **1** shows, optimal full matching with calipers retain nearly all of the data. In addition to using calipers, the fourth optimal matching in the table "accounts for clusters" by exactly matching on hospital, by which the patients are clustered. The balance chi-square test statistics and p-values shown in Table **1**, from previously discussed tests, incorporate the matching structure into the testing of balance. A statistically significant p-value indicates that the balancing score property of the propensity score is not attained with the chosen model and matching arrangement. Clearly, greedy pair matching is a poor choice because much of the data is discarded and there are different distributions of the covariates for femoral versus radial access patients.

To illustrate how the chosen optimal full match improves balance in the covariates, Figure **3** compares the standardized differences between radial and femoral PCI access before and after matching. In the figure, the closer the points are to the zero, the less the imbalance in a given covariate.

Subsequent Analysis

Following the creation of matched sets, we analyzed the data accounting for this structure. To illustrate multiple methods, we used both Cochran

Table 1: Balance Results

Matching method	% observations obtained	Balance test statistic (p-value)
Greedy pair match	46.40%	583 (< 0.001)
Optimal full match, caliper = 0.005	93.80%	423 (p <0.001)
Optimal full match, caliper = 0.01	96.20%	25.1 (p=0.092)
Optimal full match, caliper=0.01 (accounting for clusters)	95.61%	6.56 (p=0.99)

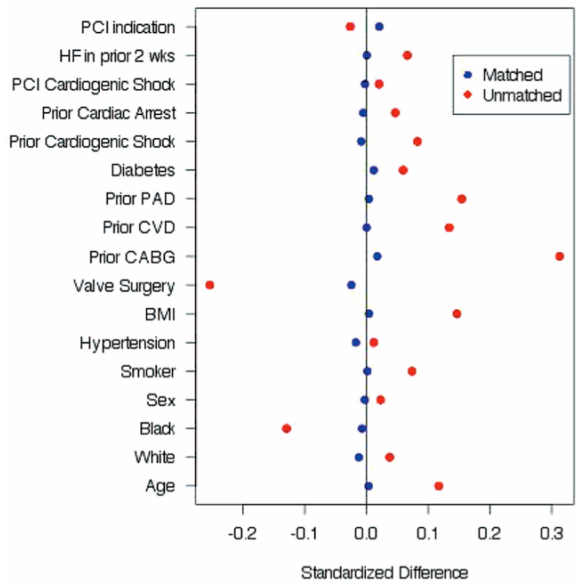

Figure 3: Standardized differences before and after matching.

Mantel Haenszel tests and conditional logistic regression. The Mantel Haenszel tests do not differ from the conditional logistic regression results presented in their statistical significance at a 5% level, with the single exception of dialysis need (p=0.157 with the Mantel Haenszel tests as compared to 0.055 with logistic regression). While not presented, it should be noted that the results from t-tests on the set of observations that were matched in the greedy pair matching do not differ substantively from the z-statistics from the conditional logistic regression with the greedy matching. As mentioned previously, this analytic strategy is insufficient as it fails to account for the matched structure of the data. At a significance level of 5%, most of the outcomes studied showed highly statistically significant differences between patients who received PCI through femoral access as opposed to radial access when the t-tests on the matched observations were performed. Yet, it is widely accepted that these outcomes should not be different across access sites, particularly to this degree. When conditional logistic regression on the optimally full matched data is used, accounting for the matched structure, the results are quite different. As Table **2** shows, at a significance level of 5%, only one of the seven outcomes besides bleeding – cardiogenic shock – showed a statistically significant difference across access sites. As can be seen in our Tables **3a** and **3b** sensitivity analysis, this finding is much less robust than the statistically significant difference in bleeding across access sites.

DISCUSSION

Propensity score matching is a useful tool to analyze observational data provided its execution respects certain standards. This comment outlines

Table 2: Outcome Results

			Greedy Match		Optimal full match with multilevel propensity score			
	Unweighted rates from matched sample		OR	z-statistic* (p-value)	Weighted rates accounting for matched sets		OR	z-statistic* (p-value)
Post PCI outcome	Femoral	Radial			Femoral	Radial		
Bleeding event	0.054	0.011	5.77	10.74 (p < 0.01)	0.031	0.015	2.73	6.23 (p < 0.01)
Dialysis need	0.012	0.002	7.13	5.20 (p < 0.01)	0.004	0.002	2.17	1.92 (p=0.055)
Cardiogenic shock	0.056	0.010	6.23	11.14 (p < 0.01)	0.022	0.012	1.98	3.97 (p< 0.01)
Heart Failure	0.054	0.017	3.41	9.35 (p < 0.01)	0.024	0.020	1.23	1.49 (p=0.14)
CVA	0.008	0.003	2.86	3.38 (p < 0.01)	0.003	0.003	1.17	0.48 (p=0.63)
Myocardial infarction	0.018	0.017	1.09	0.54 (p=0.59)	0.020	0.016	1.28	1.66 (p=0.10)

*z-statistic computed from conditional logistic regression of indicator of femoral access on indicated post PCI outcome.

Table 3a: Sensitivity Analysis for Post-PCI Bleeding

Effect Size	Prevalence in Control	Prevalence in Treatment	Adjusted 95% CI
2	0	0.25	(1.59, 2.99)
3	0	0.5	(1.00, 1.87)
3	0.2	0.25	(1.86, 3.49)
3	0	0.25	(1.33, 2.49)
4	0	0.25	(1.14, 2.14)
4	0.2	0.25	(1.82, 3.42)
4	0	0.5	(0.80, 1.50)

Statistically significant (5% level) difference between radial and femoral maintained.
Statistically significant (5% level) difference between radial and femoral lost.

Table 3b: Sensitivity Analysis for Post-PCI Cardiogenic Shock

Effect Size	Prevalence in Control	Prevalence in Treatment	Adjusted 95% CI
2	0	0.25	(1.13, 2.22)
3	0	0.5	(0.71, 1.39)
3	0.2	0.25	(1.32, 2.59)
3	0	0.25	(0.94, 1.85)
4	0	0.25	(0.81, 1.58)
4	0.2	0.25	(1.29, 2.53)
4	0	0.5	(0.57, 1.11)

Statistically significant (5% level) difference between radial and femoral maintained.
Statistically significant (5% level) difference between radial and femoral lost.

eight traits of a good propensity score matched analysis that are essential for the proper use of this valuable analytic technique in a clinical context. Two mistakes are commonly made when employing propensity score matching techniques. First, proper balance assessments and sensitivity analyses are often skipped. Second, balance assessments, if included, and outcome analyses after matching fail to incorporate the matching structure, not recognizing that the aim of propensity score matching is to approximate a block-randomized study. If the eight traits outlined here are incorporated into a matched study, these two mistakes will not be made. The propensity score model will be correctly estimated, proper balance assessments will be performed, the analysis will incorporate the matched structure, and sensitivity analyses will assess the robustness of the findings, as illustrated in our case study of the impact of PCI access site location. Propensity score matching must be performed carefully to leverage its power to better use observational data to answer clinical questions.

Summary: 8 Traits of a Good Propensity Score Matched Analysis

1. The population should be carefully selected to ensure that there would be no clinical factor that would account for marked differences in, or total absence of exposure to one or more of the treatments under investigation.

2. The propensity score model should include only pretreatment variables.

3. The propensity score specification can be assessed following the creation of matched sets on the basis of balance in the covariates as opposed to using standard regression diagnostics.

4. Assessments of a chosen matching structure should utilize balance tests and not consider outcome analyses.

5. Evaluations of balance should incorporate the matching structure and not outcome information.

6. Subsequent outcome analyses must account for the matching structure; they cannot simply analyze the matched sample.

7. Sensitivity analysis is necessary to assess unmeasured confounding.

8. When evaluating findings, always pay attention to biological plausibility; results that violate plausibility may indicate the presence of residual confounding that persists after statistical adjustment.

FUNDING SOURCE

None.

DISCLOSURES

Hitinder S. Gurm receives research funding from Blue Cross Blue Shield of Michigan and the National Institute of Health. None of the authors have any conflicts relevant to this study.

REFERENCES

[1] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70: 41-55. http://dx.doi.org/10.1093/biomet/70.1.41

[2] Luo Z, Gardiner JC, Bradley CJ. Applying propensity score methods in medical research: pitfalls and prospects. Medical Care Research and Review 2010; 67: 528-554. http://dx.doi.org/10.1177/1077558710361486

[3] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. American journal of epidemiology 2006; 163: 1149-1156. http://dx.doi.org/10.1093/aje/kwj149

[4] Rubin DB, Thomas N. Matching using estimated propensity scores: relating theory to practice. Biometrics 1996: 249-264. http://dx.doi.org/10.2307/2533160

[5] Rosenbaum PR, SpringerLink (Online service). Design of observational studies. New York: Springer, 2010: 1 online resource (xviii, 384 p.

[6] Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Statistics in medicine 2007; 26: 734-753. http://dx.doi.org/10.1002/sim.2580

[7] Griswold ME, Localio AR, Mulrow C. Propensity score adjustment with multilevel data: setting your sites on decreasing selection bias. Annals of internal medicine 2010; 152: 393-395. http://dx.doi.org/10.7326/0003-4819-152-6-201003160-00010

[8] Arpino B, Mealli F. The specification of the propensity score in multilevel observational studies. Computational Statistics & Data Analysis 2011; 55: 1770-1780. http://dx.doi.org/10.1016/j.csda.2010.11.008

[9] Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. Statistics in medicine 2007; 26: 20-36. http://dx.doi.org/10.1002/sim.2739

[10] Hansen BB, Klopfer SO. Optimal full matching and related designs *via* network flows. Journal of Computational and Graphical Statistics 2006; 15.

[11] Gu XS, Rosenbaum PR. Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics 1993; 2: 405-420. http://dx.doi.org/10.2307/1390693

[12] Ming K, Rosenbaum PR. Substantial gains in bias reduction from matching with a variable number of controls. Biometrics 2000; 56: 118-124. http://dx.doi.org/10.1111/j.0006-341X.2000.00118.x

[13] Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. Journal of the royal statistical society: series A (statistics in society) 2008; 171: 481-502. http://dx.doi.org/10.1111/j.1467-985X.2007.00527.x

[14] Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Medical Decision Making 2009. http://dx.doi.org/10.1177/0272989X09341755

[15] Hansen BB, Bowers J. Covariate balance in simple, stratified and clustered comparative studies. Statistical Science 2008: 219-236. http://dx.doi.org/10.1214/08-STS254

[16] Rosenbaum PR. Observational studies: Springer, 2002. http://dx.doi.org/10.1007/978-1-4757-3692-2

[17] Hosman CA, Hansen BB, Holland PW. The sensitivity of linear regression coefficients' confidence limits to the omission of a confounder. The Annals of Applied Statistics 2010; 4: 849-870. http://dx.doi.org/10.1214/09-AOAS315

[18] Lin D, Psaty BM, Kronmal R. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998: 948-963. http://dx.doi.org/10.2307/2533848

[19] Gurm HS, Smith DE, Collins JS *et al*. The relative safety and efficacy of abciximab and eptifibatide in patients undergoing primary percutaneous coronary intervention: insights from a large regional registry of contemporary percutaneous coronary intervention. Journal of the American College of Cardiology 2008; 51: 529-35. http://dx.doi.org/10.1016/j.jacc.2007.09.053

[20] Moscucci M, Rogers EK, Montoye C *et al*. Association of a continuous quality improvement initiative with practice and outcome variations of contemporary percutaneous coronary interventions. Circulation 2006; 113: 814-22. http://dx.doi.org/10.1161/CIRCULATIONAHA.105.541995

[21] Kline-Rogers E, Share D, Bondie D *et al*. Development of a multicenter interventional cardiology database: the Blue Cross Blue Shield of Michigan Cardiovascular Consortium (BMC2) experience. Journal of interventional cardiology 2002; 15: 387-92. http://dx.doi.org/10.1111/j.1540-8183.2002.tb01072.x

Received on 16-04-2015 Accepted on 30-04-2015 Published on 21-05-2015

http://dx.doi.org/10.6000/1929-6029.2015.04.02.7