

REDUCING BIAS IN A RETROSPECTIVE CASE-CONTROL STUDY: AN APPLICATION OF PROPENSITY SCORE MATCHING

Alex Exuzides¹, Chris Colby¹, Jonathan Goldman², Anne Waaler³

1. ICON Clinical Research, San Francisco, CA; 2. ICON Medical Imaging, San Francisco, CA; 3. GE Healthcare, Horten, Norway



A Symbol of Excellence

Introduction

- In retrospective studies, where treatment selection is non-random, cases and controls frequently show large imbalances in patient characteristics.
- These important differences in patient characteristics can create substantial bias in treatment comparisons.
- In particular, patients with advanced or unstable disease may receive the most intensive treatment.
- These differences must be adjusted for in order to reduce selection bias and accurately estimate the treatment effect.
- While it is possible to match on a small number of key patient characteristics, if there are numerous covariates, exact matches may not be possible.
- One method to reduce treatment selection bias is the use of propensity scores. The propensity score utilizes all available covariates to estimate the probability of receiving the treatment.
- Once propensity scores are estimated, cases can be matched to one or more individual controls based on the propensity scores.
- This study provides results from the application of propensity score matching that utilizes the nearest-neighbor technique.

Methods

Data Source and Sample Selection

- This study utilized Premier's Perspective™ Comparative Database, the largest hospital service-level database in the U.S. derived from detailed hospital discharge data.
- All adult patients undergoing inpatient echocardiography between Jan. 2003 and Oct. 2005 were identified (n=2,588,722).
- Of interest is the use of an echocardiography contrast agent with perflutren protein-type A microspheres injectable suspension, USP (contrast agent).
- Out of all echocardiography patients in the database, a total of 22,499 received the contrast agent.
- Echocardiography patients receiving other contrast agents were excluded from the analysis.
- From the 22,499 contrast agent patients, 2,900 had diagnoses meeting the criteria for critical illness (heart failure, acute myocardial infarction, arrhythmia, respiratory failure, pulmonary embolism, emphysema, and pulmonary hypertension).

Table 1. Baseline Characteristics by Contrast Agent Use Before Matching

Variables	Contrast Agent Patients (N=2,900)	Non-Contrast Agent Patients (N=208,878)	P-value ¹
	Mean or %	Mean or %	
Age	65.6	69.4	<0.001
Gender (%Female)	38.62%	49.80%	<0.001
Race (%White)	73.45%	65.65%	<0.001
Geographic region: Midwest	30.66%	19.20%	<0.001
Geographic region: Northeast	15.76%	19.43%	<0.001
Geographic region: South	50.97%	49.62%	0.15
Geographic region: West	2.62%	11.75%	<0.001
Urban/Rural location (%Urban)	91.83%	89.89%	<0.001
Hospital type (%Teaching)	63.59%	45.96%	<0.001
Admission through ER	57.38%	70.54%	<0.001
Level of care: Mechanical Ventilation	17.00%	11.02%	<0.001
Level of care: ICU ²	22.14%	17.20%	<0.001
Level of care: CCU ³	15.00%	13.00%	0.002
Anticoagulant treatment	95.90%	92.26%	<0.001
Concomitant medication usage	11.07%	8.77%	<0.001
Deyo-modified CCI ⁴ (DM-CCI)	2.45	2.25	<0.001
Cerebrovascular disease (part of DM-CCI)	9.41%	7.55%	<0.001
Chronic pulmonary disease (part of DM-CCI)	34.93%	31.54%	<0.001
Rheumatologic disease (part of DM-CCI)	1.55%	2.25%	0.01
Liver disease (part of DM-CCI)	1.03%	0.91%	0.49
Diabetes Type 1 (part of DM-CCI)	35.24%	29.17%	<0.001
Diabetes Type 2 (part of DM-CCI)	17.72%	11.76%	<0.001
Renal disease (part of DM-CCI)	9.10%	6.30%	<0.001
Malignancy (part of DM-CCI)	5.72%	7.86%	0.003
Severe Liver Disease (part of DM-CCI)	2.07%	1.13%	0.007
Metastatic Cancer (part of DM-CCI)	4.76%	8.38%	0.006
CAD ⁵ comorbidity	61.59%	54.85%	<0.001
Muscular-Skeletal comorbidity	2.17%	2.13%	0.89
Depression comorbidity	5.86%	6.65%	0.09

¹ P-value using the Wilcoxon test for continuous variables and the Mantel-Haenszel test for categorical variables.
² In Intensive Care Unit, other than those receiving Mechanical Ventilation, at any point in hospital admission up to the same day of echocardiography.
³ In Coronary Care Unit at any point in hospital admission up to the same day of echocardiography.
⁴ Charlson Comorbidity Index.
⁵ Coronary Artery Disease.

Statistical Methods

- The first step in the propensity score matching was to model the choice between contrast agent vs. non-contrast agent echocardiogram. For that, we used a logistic regression model with the variables listed in Table 1.
- A stepwise selection process was implemented to choose those variables with the best predictive power. We used SAS® PROC LOGISTIC with the SELECTION=STEPWISE option and the default significance level of 0.05.
- For each contrast agent patient, we selected four matched control patients with similar propensity scores. The selection was based on the smallest difference between propensity scores among cases and controls. This method is commonly referred to as the nearest neighbor matching.
- We implemented the nearest neighbor matching using a SAS® macro (greedy matching algorithm¹). This matching algorithm performs up to 7 passes to find one or more matched controls for each case: First, the algorithm searches for controls with similar propensity scores within a tolerance of 0.0000001. The algorithm progressively relaxes the tolerance by one digit until it gets to 0.1.
- We limited the selection of matched controls to four, because selection of more than four controls per case does not result in substantial gain in statistical efficiency.²
- This implementation of propensity score matching was performed using SAS® 9.1.3.^{3,4}

Results

Population characteristics

- Baseline patient characteristics by contrast agent use are shown in Table 1.
- Nearly all of the baseline characteristics showed statistically significant differences between cases and controls at the 0.01 level (Table 1), indicating the potential of serious selection bias in treatment choice.
- Specifically, patients receiving the contrast agent were more likely to have cerebrovascular, renal, and diabetic comorbidities, and were more likely to be treated in the ICU or receive mechanical ventilation (Table 1).

Results (Cont'd)

Table 2. Stepwise Regression Results

Effect	Wald Chi-Square	P-value ¹
Gender (%Female)	64.4	<0.001
Admission through ER	119.8	<0.001
Level of Care (MV/ICU/CCU/Other)	137.0	<0.001
Age	161.6	<0.001
Region (Midwest/Northeast/South/West)	256.8	<0.001
Anticoagulant treatment	16.3	<0.001
Concomitant medication usage	6.5	0.01
Hospital type (Teaching/Non-teaching)	152.7	<0.001
Race (White/Nonwhite)	68.1	<0.001
Cerebrovascular disease (part of DM-CCI)	7.7	0.006
Chronic pulmonary disease (part of DM-CCI)	7.6	0.006
Diabetes Type 1 (part of DM-CCI)	71.2	<0.001
Diabetes Type 2 (part of DM-CCI)	52.6	<0.001
Renal disease (part of DM-CCI)	15.1	<0.001
Malignancy (part of DM-CCI)	6.3	0.01
Severe Liver Disease (part of DM-CCI)	3.9	0.05
CAD comorbidity	5.6	0.02
Depression comorbidity	4.4	0.04

Propensity score matching

- Table 2 lists the variables selected by the stepwise selection process for the logistic regression that models the choice between contrast agent vs. non-contrast agent echocardiogram.
- By matching on the derived propensity scores, the greedy matching algorithm identified 4 matched controls for each of the 2,900 cases.
- Achieving a 100% matching rate reduces the bias that can occur when the sickest (or healthiest) patients of one treatment group are eliminated from the analysis.
- After implementing the propensity score matching algorithm, differences in patient characteristics largely disappeared and the two treatment groups appeared to be balanced (Table 3).
- In particular, differences between the two groups became insignificant for the all of the comorbid disease categories, as well as the level of care variables (Table 3).
- After matching, only one variable remained statistically significant: Higher concomitant medication usage among cases; P-value=0.006 (Table 3).
- After matching, only one variable remained statistically significant: Higher concomitant medication usage among cases; P-value=0.006 (Table 3).

Table 3. Baseline Characteristics by Contrast Agent Use After Matching

Variables	Contrast Agent Patients (N=2,900)	Non-Contrast Agent Patients (N=11,600)	P-value ¹
	Mean or %	Mean or %	
Age	65.6	66.0	0.24
Gender (%Female)	38.62%	37.61%	0.32
Race (%White)	73.45%	73.93%	0.60
Geographic region: Midwest	30.66%	30.06%	0.53
Geographic region: Northeast	15.76%	13.63%	0.003
Geographic region: South	50.97%	53.47%	0.02
Geographic region: West	2.62%	2.84%	0.53
Urban/Rural location (%Urban)	91.83%	91.22%	0.30
Hospital type (%Teaching)	63.59%	62.69%	0.37
Admission through ER	57.38%	57.63%	0.81
Level of care: Mechanical Ventilation	17.00%	17.11%	0.89
Level of care: ICU ²	22.14%	22.09%	0.95
Level of care: CCU ³	15.00%	14.40%	0.41
Anticoagulant treatment	95.90%	96.82%	0.013
Concomitant medication usage	11.07%	9.37%	0.006
Deyo-modified CCI ⁴ (DM-CCI)	2.45	2.42	0.29
Cerebrovascular disease (part of DM-CCI)	9.41%	8.62%	0.18
Chronic pulmonary disease (part of DM-CCI)	34.93%	35.15%	0.83
Rheumatologic disease (part of DM-CCI)	1.55%	2.00%	0.11
Liver disease (part of DM-CCI)	1.03%	1.12%	0.69
Diabetes Type 1 (part of DM-CCI)	35.24%	36.53%	0.20
Diabetes Type 2 (part of DM-CCI)	17.72%	16.48%	0.28
Renal disease (part of DM-CCI)	9.10%	7.48%	0.04
Malignancy (part of DM-CCI)	5.72%	5.03%	0.30
Severe Liver Disease (part of DM-CCI)	2.07%	1.53%	0.24
Metastatic Cancer (part of DM-CCI)	4.76%	6.05%	0.29
CAD ⁵ comorbidity	61.59%	62.45%	0.39
Muscular-Skeletal comorbidity	2.17%	1.76%	0.14
Depression comorbidity	5.86%	5.11%	0.11

¹ P-value using the Wilcoxon test for continuous variables and the Mantel-Haenszel test for categorical variables.
² In Intensive Care Unit, other than those receiving Mechanical Ventilation, at any point in hospital admission up to the same day of echocardiography.
³ In Coronary Care Unit at any point in hospital admission up to the same day of echocardiography.
⁴ Charlson Comorbidity Index.
⁵ Coronary Artery Disease.

Limitations

- Propensity score matching can only control for observed covariates. Any unobserved patient characteristics that affect both treatment selection and outcomes may still create selection bias.
- Despite the large differences in patient characteristics, this patient population had sufficient overlap between cases and controls to match every case with 4 controls. Not all study populations will yield complete matches.
- When a large number of cases are eliminated, because the algorithm fails to find a control with a similar propensity score, the validity of the resulting analysis may be questionable.

Conclusion

- Propensity score matching is a useful way to reduce selection bias in retrospective studies, especially when the treatment choice may be affected by a large number of covariates.
- Propensity score matching can be easily implemented in SAS® using the greedy matching algorithm.

References

- Rosenbaum PR and Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician. 1985; 39(1): 33-8.
- Parson LS. Performing a 1:N case-control match on propensity score. Proceeding of the 29th Annual SAS® Users Group International Conference; SAS® Institute Inc.; 2004.
- Ury HK. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. Biometrics. 1975 Sep; 31(3): 643-9.
- SAS/STAT software, Version 9.1.3 of the SAS® System for Windows. Copyright 2003 SAS® Institute, Inc.

Presented at the ISPOR 12th Annual European Congress; Paris, France, October 2009