

Pseudo-Randomization In Retrospective Analysis Using The Generalized Multinomial Logit For Propensity Score Generation

SM Wilson; T Mayne
DaVita Inc., Denver, CO, USA

Introduction

- Pseudo-randomization via propensity score matching (PSM) has been used to help mitigate channeling bias in retrospective studies for more than 2 decades.¹
- Often this method of correction involves the calculation of inverse probability weights, which can seem like a "black box."
- Imbens has reported that a causal interpretation of treatment effects can be obtained by using the generalized propensity score (PS) for any number of treatments and averaging overall treatment possibilities.²
- Using the PS to select comparable groups of subjects is appealing to analyze and present data in a widely recognizable format.

Objective

 To develop and test a three-way PSM algorithm to provide pseudo-randomization of subjects into 3 groups to allow for comparable groups in a retrospective study.

Methods

- An empirical test using three treatment groups predicted from patient gender, age, race, and Charlson Comorbidities Index score was conducted to see if balance on these variables could be obtained using multi-way PSM.
- Logistic regression using the generalized multinomial logit linking function was used to calculate estimates of the PS:
- The probability of having received 3 putatively interchangeable drugs from demographics (race, gender, age) and comorbidities (Charlson Comorbidities Index) in a large, retrospective database.
- The most costly drug was used as a reference group and the probability of each treatment group having received the reference drug was retained as the PS.

Methods (continued)

- In the initial analysis, there were 3 treatment groups with 23,912 individuals, 4,789 individuals, and 4,318 individuals, respectively.
- Random subsets of 1/4 and 1/10 the original sample were constructed for the purpose of assessing multi-group PSM effectiveness in constructing comparable groups via pseudo-randomization with varying starting sample sizes.
- PSM was conducted using calipers ranging from 8 digits to one digit of PS.

Results

Table 1: Pre- and Post-PSM by Age, Charlson Score, and Gender

	Pre-PSM		Post-PSM	
Groups	Mean (SD) N (%)	Omnibus <i>p</i> -value	Mean (SD) N (%)	Omnibus p-value
Age				
Group A	56.4 (14.4)		57.8 (11.7)	
Group B	52.5 (14.0)	< 0.0001	57.7 (11.8)	0.84
Group C	55.0 (14.6)		57.9 (11.8)	
Charlson Score				
Group A	5.1 (2.1)		5.1 (2.0)	
Group B	4.7 (2.0)	< 0.0001	5.2 (2.0)	0.63
Group C	5.0 (2.1)		5.1 (1.9)	
Gender (% Females)				
Group A	1947 (45.1)		1814 (53.7)	
Group B	1947 (40.7)	< 0.0001	1810 (53.5)	0.99
Group C	11930 (49.9)		1817 (53.7)	

Results (continued)

- For all sample sizes, prior to PSM, significant differences existed among the three regimens for all variables: gender, race, age, and comorbidities (all *p*-values < 0.0001).
- Following PSM, there were 3,381 matched triplets in the full sample.
- There were no significant differences among groups for gender, age, or comorbidities.
- There were significant, but small differences that remained for racial representation.
- In the smaller samples, 966 and 416 matched triplets were retained.
- There were no significant differences for any variables.
- Assessment of among-group differences before and after PSM were conducted using Chi-square tests for categorical variables and GLM analysis, with different scores and their confidence intervals for continuous variables.

Discussion

- Pseudo-randomization via PSM with the generalized multinominal logit appears to balance groups for demographic variables.
- These groups appear similar as if they had been randomized a priori. Thus, pseudo-randomization appears to have occurred.
- These groups are representative of the patient population as a whole. Thus, the breadth of the PS distribution appears to have been adequate in all three groups.
- Additional research is required to assess the efficacy of this methodology on a broader set of pre-treatment predictors, potentially including baseline biomarkers.
- A critical check is to confirm that representativeness on demographic variables also provides representativeness (and comparability) on variables likely related to channeling bias.

Conclusions

- Logistic regression using a generalized multinomial logit link appears to provide a good PS from which pseudo-randomization into three groups can be performed in a retrospective sample.
- The groups were comparable in PS range with nearly complete overlap.
- Prior to PSM, the three groups were significantly different in age, gender, race, and disease burden (all *p*-values < 0.0001).
- -Following PSM, the three groups were comparable on each measure.

References

- 1. Rosenbaum P, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Statist Assoc.* 1984;79:516–524.
- 2. Imbens G. The role of the propensity score in estimating dose-response functions. *Biometrika*. 2000;87:706–710.

Acknowledgments

Our sincere appreciation to the teammates in more than 1600 DaVita clinics who work every day to take care of patients but also to ensure the extensive data collection on which our work is based. We thank DaVita Clinical Research® (DCR®), and specifically acknowledge Barb Nambu, PhD of DCR for editorial contributions in preparing this poster. DCR is committed to advancing the knowledge and practice of kidney care.

*Correspondence: Steve.Wilson@DaVita.com

Poster available at www.davitaclinicalresearch.com/directory.asp

International Society for Pharmacoeconomics and Outcomes Research 17th Annual International Meeting, Washington, D.C., June 2–6, 2012