Technical Report 53
Multiple Treatments and Propensity Scores

Rodney Sparapani[1*] and Prakash Laud[1,2]
[1]Center for Patient Care and Outcomes Research, [2]Division of Biostatistics
Medical College of Wisconsin, Milwaukee, WI 53226
[*]Corresponding Author: rsparapa@mcw.edu

December 30, 2005

1 Introduction

Propensity score adjustment is a statistical method that attempts to control for selection bias in observational studies. The methodology was described in two papers by Rosenbaum and Rubin (1983,1984). There are many applications of this method when with two treatments as discussed, for example, in D'Agostino (1998) and the references there. For three or more treatments the literature is sparse, especially with regard to practical recommendations for implementation. We begin by describing the methodology as employed for two treatments.

Suppose that we have an outcome, Y, and two treatments, $Z = 0$ and $Z = 1$. In a randomized study, treatments are randomized to patients so that the two groups can be assumed to be statistically balanced with respect to any covariates, measured or unmeasured. In an observational study, such balance cannot be expected and patients' treatment assignment (or, more appropriately, choice) is typically related to covariates, $\boldsymbol{X}$ (where **bold** represents a vector). Under the circumstances, we can estimate the treatment effect as

$$\delta_{\boldsymbol{X}=\boldsymbol{a}} = E[Y|Z = 1,\ \boldsymbol{X} = \boldsymbol{a}] - E[Y|Z = 0,\ \boldsymbol{X} = \boldsymbol{a}].$$

However, δ is a function of $\boldsymbol{X}$ which limits its use as a measure of effectiveness in most circumstances. One can, however, search for a balancing score, $b(\boldsymbol{X})$, which is a function that accepts the covariate vector as its argument, and results in a scalar such that patients with similar balancing scores have similar covariate values. The treatment effect can be now taken to be

$$\delta_{b(\boldsymbol{X})=c} = E[Y|Z = 1,\ b(\boldsymbol{X}) = c] - E[Y|Z = 0,\ b(\boldsymbol{X}) = c].$$

Although still a function of the balancing score, δ depends on a scalar rather than a vector achieving, at the very least a reduction in dimension. Further, if there are ranges of values which form homogeneous strata with respect to the distribution of $\boldsymbol{X}$, we can consider the treatment effect as constant within strata. For this purpose, let

$$\delta_{c\leq b(\boldsymbol{X})<d} = E[Y|Z = 1,\ c \leq b(\boldsymbol{X}) < d] - E[Y|Z = 0,\ c \leq b(\boldsymbol{X}) < d].$$

A choice of $b(\boldsymbol{X})$ is provided by Theorems 1 and 2 of Rosenbaum and Rubin (1983). These authours have shown that the propensity score, defined by $p(\boldsymbol{X}) = P[Z = 1|\boldsymbol{X}]$, is a balancing score as defined above. In practice, this propensity score can be estimated by logistic regression:

$$\log \frac{p(\boldsymbol{X})}{1 - p(\boldsymbol{X})} = \boldsymbol{\beta}' \boldsymbol{X}.$$

Groups of like propensity scores can be made by suitable cutoffs, leading to

$$\delta_j = E[Y|Z = 1,\ p_{j-1} \leq p(\boldsymbol{X}) < p_j] - E[Y|Z = 0,\ p_{j-1} \leq p(\boldsymbol{X}) < p_j]$$

where p_j are typically taken to be quintiles of the estimated propensity scores with $p_0 = 0,\ p_5 = 1$. If the δ_j are similar, then their average is taken as the estimate of the overall δ. Otherwise, more advanced strategies need to be employed, such as regression. The scope of this report is limited to the construction of groups with like propensity scores when there are three ordinal treatments.

2 Motivating Scenario

This exploration is motivated by an observational study of breast cancer survivors, each receiving one of three treatments (Gilligan et al.,2005). The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute is an authoritative source of information on cancer incidence and survival in the United States. Data is collected on all patients suffering all cancers (except non-melanoma skin cancer) from 1973-99 in the following registry areas: 5 states (CT, HI, IA, NM, UT) and 6 metropolitan areas (Atlanta, Detroit, Los Angeles*, San Francisco/Oakland, San Jose/Monterey*, Seattle/Puget Sound) * 1988-99. The SEER-Medicare Database is created by linking Medicare identifiers to SEER patients aged 65+ (found for about 94%) and all claims collected including hospital, physician and clinic.

A volume cohort was constructed consisting of 22,271 female SEER-Medicare patients, aged 65+, suffering an in situ and/or invasive breast cancer during 1994-96 with surgical treatment performed at 457 hospitals where the patient and the hospital were in the same SEER registry area. A subset of the volume cohort called the study cohort was constructed consisting of 11,187 female SEER-Medicare patients, aged 66-95, suffering their first invasive cancer which was a unilateral, Stage I/II, microscopically confirmed breast cancer during 1994-96 who were continuously enrolled in non-HMO Medicare from one year prior to their breast cancer diagnosis to four months afterward or until death if sooner.

For each patient, their hospital of care was identified and its associated volume computed (determined by the annual average number of patients from the volume cohort who received their surgical treatment at that hospital). In this exercise, hospital volume is considered to be the treatment. Three categories of treatment are considered to be clinically meaningful based on average annual hospital volume: low, medium and high. For these labels to be meaningful, it was deemed desirable that the high volume consist of at least double the annual volume considered low. In addition, approximately equal number of cases in each group was also deemed desirable. The groups made with annual volumes of $(0,\ 20)$, $[20,\ 40)$ and $40+$ satisfied these criteria.

It cannot be assumed that each patient was randomly assigned to their hospital. On the contrary, there are many factors that contribute to where a patient receives their breast cancer surgery such as their breast cancer prognosis, prospective treatment, age, health, race, socio-economic status, locality, etc. Ignoring the relationship between these factors and their treatment assignment could invalidate the analysis. So, these important covariates must be handled appropriately. If this were a two-treatment study, then a propensity score adjustment could be readily employed as has been discussed in the literature cited above. However,

since this is a three-treatment study, a generalization of the two-treatment propensity score adjustment is necessary.

3 Proposed Method and Subsequent Results

The broad outlines of an extension to three or more treatments are contained in Joffe and Rosenbaum (1999), without details of implementation. For the purpose of this report, we assume three ordered treatments, $Z \in \{0, 1, 2\}$. The propensity scores for three treatments can be denoted as

$$p_1(\boldsymbol{X}) = P[Z = 1|\boldsymbol{X}],\ p_2(\boldsymbol{X}) = P[Z = 2|\boldsymbol{X}] \text{ and } p_0(\boldsymbol{X}) = 1 - p_1(\boldsymbol{X}) - p_2(\boldsymbol{X}).$$

Ordered treatment propensity scores can be estimated, for example, by continuation ratio logistic regression:

$$\log \frac{p_2(\boldsymbol{X})}{1 - p_2(\boldsymbol{X})} = \boldsymbol{\beta_2'X}$$

and

$$\log \frac{P[Z = 1|Z < 2, \boldsymbol{X}]}{1 - P[Z = 1|Z < 2, \boldsymbol{X}]} = \boldsymbol{\beta_1'X}.$$

A proposed method to construct propensity score groups from these estimated probabilities is to graph the estimated probabilities such that p_0 is on the horizontal axis and p_2 on the vertical axis. It appears natural then to construct groups by drawing lines of unit slope that break up the total sample into a desired number of equal sized groups. Algebraically, it is easy to see that this amounts to ordering observations according to the values of $p_1 + 2p_2$ and making equal quantile groups.

In the results that follow, three groups are used since they have an attractive interpretation with respect to volume. For example, the propensity score group with the smallest third of $\hat{p_1} + 2\hat{p_2}$ valuess can be thought of as the third of the cohort who would have received treatment at a low volume hospital based on their covariates.

The SEER-Medicare Database contains the following relevant demographic and clinical covariates: age, race/ethnicity, SEER registry area, ZIP code per capita income (PCI) from the 1990 Census, tumor characteristics and treatment performed. In addition, from the Medicare data, a measure of the patient's non-cancer health can be constructed; Charlson/Deyo/Klabunde comorbidity index, a weighted count of diagnosis groups found in Medicare claims in the year prior to breast cancer: 0, 1, 2+. Table 1 shows these characteristics for the cohort of 22,271 patients, broken down by the annual volume group of the hospital where the patient underwent the surgery. (Note here the following acronym definitions: ER, estrogen receptor; PR, progesterone receptor; LND, lymph node dissection; BCS, breast conserving surgery.) The table clearly shows the selection bias. For example, higher volume hospitals treat patients with higher incomes and lower comorbidities.

To account for this selection bias, a continuation ratio logistic regression was fitted to the observed volume group category. The following covariates were found to be significant predictors of hospital volume (in order of importance): MSA, ZIP PCI, race, ER/PR status known/unknown, comorbidity, lymph node status known/unknown and tumor grade known/unknown. The estimated probabilities of medium and high volume groups are plotted in Figure 1. In the graph on the left, the observed treatment assignment is shown by the color used. The graph on the right side shows the predicted treatment assignment (or the planar tertile propensity groups) based on the method proposed in this report. Notice that there are more patients being treated at a low volume hospital, indicated by blue dots, in the lower right where the probability of low volume is high and the probability of high volume is low, etc.

Table 1

Volume Groups Annual, Average	Low >0-	Med. 20-	High 40+
Patients (%)	32.0	33.0	35.0
Hospitals (%)	71.6	19.0	9.4
White (%)	89.5	92.7	90.5
Black (%)	5.4	3.6	7.3
Hispanic (%)	4.3	1.9	0.9
Median ZIP PCI ($)	13352	16368	17770
Node Unk. (%)	8.9	7.2	11.1
Grade Unk. (%)	21.4	19.0	16.8
ER/PR + (%)	65.8	71.8	69.3
Comorbidity 0 (%)	71.3	75.8	76.5
Mastectomy+LND (%)	53.5	48.4	43.5
BCS+LND (%)	26.0	30.4	32.1
Radiotherapy (%)	68.1	75.6	74.8

Figure 1

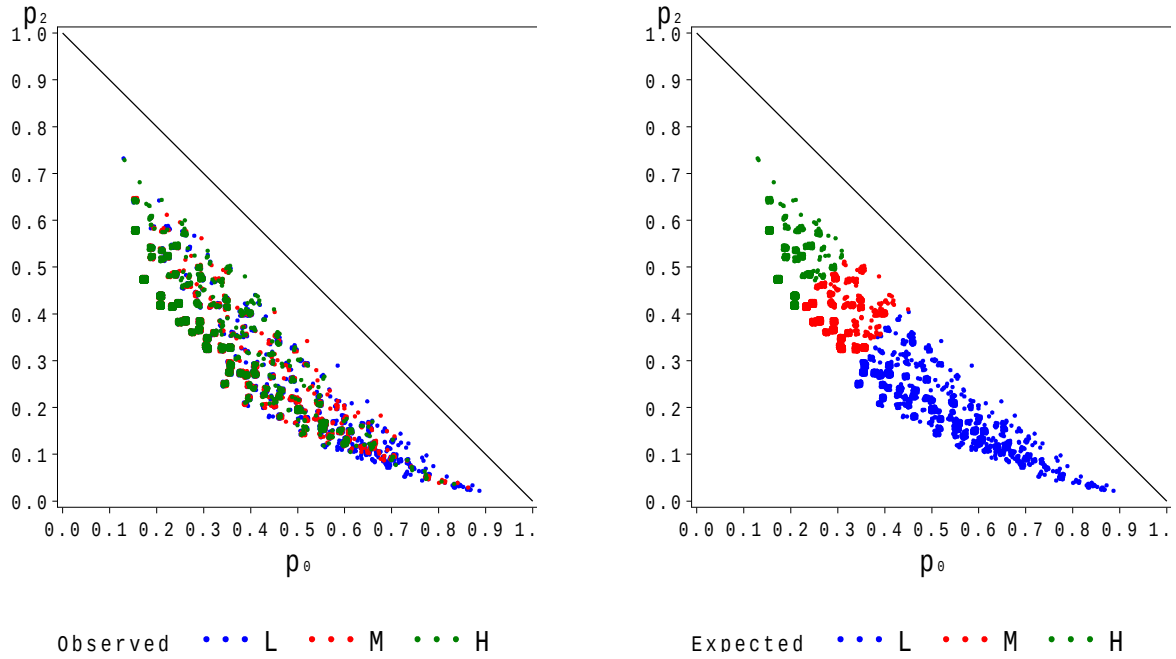

Table 2

Volume Groups	Observed		Assigned Low		Assigned High	
	Low	High	Low	High	Low	High
Total	3596	3931	1886	730	778	1990
White	89.5	90.5	89.0	86.4	92.2	93.8
Black	5.4	7.3	2.0	4.7	6.9	5.4
Hispanic	4.3	0.9	8.3	4.8	0.0	0.0
Median ZIP PCI	13352	17770	11643	12294	19315	20043
Node Unk.	8.9	11.1	8.5	7.0	11.6	15.2
Grade Unk.	21.4	16.8	25.6	25.8	13.0	12.4
ER/PR +	65.8	69.3	64.5	64.4	81.5	81.9
Comorbidity 0	71.3	76.5	67.7	67.7	79.9	83.9
Mastectomy+LND	53.5	43.5	60.9	56.7	40.7	37.2
BCS+LND	26.0	32.1	21.1	24.1	34.4	35.5
Radiotherapy	60.2	68.8	61.0	68.8	74.4	76.9

The results of how the proposed method has addressed the balancing of covariates within propensity are shown in Table 2. To conserve space, only the low and high assigned and observed groups are shown. The first (double) column repeats information from Table 1 for comparison purposes. It can be seen that in each of the last two columns, the two columns tiled Low and High are more homogeneous on the covariates than in the column titled "Observed"; and this is even true of covariates which were not independent variables in the propensity scores model. Thus the differences in covariates are seen to be much more prominent between the last two columns than within them indicating a more balanced covariate distribution across observed treatments within each assigned group assigned via the proposed propensity method.

4 References

1. Gilligan MA, Neuner J, Zhang X, Sparapani R, Laud PW, Nattinger AB. Relationship of Hospital Volume of Breast Cancer Operations and Five-Year Survival after Treatment for Early Stage Breast Cancer. Submitted, 2005.

2. PR Rosenbaum and DB Rubin. The central role of the propensity score in observational studies for causal effects. Biometrika (1983), 70(1), pp. 41-55.

3. PR Rosenbaum and DB Rubin. Reducing bias in observational studies using subclassification on the propensity score. JASA (1984), 79(387), pp. 516-524

4. RB D'Agostino, Jr. Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat in Med (1998), 17, pp. 2265-2281.

5. MM Joffe and PR Rosenbaum. Invited commentary: Propensity scores. AJE (1999), 150(4), pp. 327-333