Paper 11420-2016

Integrating SAS® and R to Perform Optimal Propensity Score Matching

Lucy D'Agostino McGowan and Robert Alan Greevy, Jr.,
Vanderbilt University, Department of Biostatistics

ABSTRACT

In studies where randomization is not possible, imbalance in baseline covariates (confounding by indication) is a fundamental concern. Propensity score matching (PSM) is a popular method to minimize this potential bias, matching individuals who received treatment to those who did not to reduce the imbalance in pre-treatment covariate distributions. PSM methods continue to advance as computing resources expand. Optimal matching, which selects the set of matches that minimizes the average difference in propensity scores between mates, has been shown to outperform less computationally intensive methods. However, many find the implementation daunting. SAS/IML® software allows the integration of optimal matching routines that execute in R, e.g. the R nbpMatching package. This paper walks through performing optimal PSM in SAS® through implementing R functions. It covers the propensity score creation in SAS, the matching procedure, and the post-matching assessment of covariate balance using SAS/STAT® 13.2 and SAS/IML procedures.

INTRODUCTION

In studies where randomization is not possible, statistical methods can be employed to control for potential bias. One method used to control this bias in observational studies is propensity score matching, where individuals who receive a treatment are matched to those who do not in order to reduce the imbalance in pre-treatment covariates (D'Agostino 1998, Rosenbaum and Rubin 1985). As computing resources expand, propensity score matching methods continue to advance. Optimal matching, which selects the set of matches that minimizes the average difference in propensity scores between mates, has been shown to outperform less computationally intensive methods (Rosenbaum 1989). SAS/IML® software allows the integration of optimal matching routines that execute in R. This allows for a single program to include both SAS® and R code, seamlessly integrating the two languages.

As a driving example, we use the Right Heart Catheterization dataset (available online at http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/rhc.html). This dataset was used to assess the effectiveness of right heart catheterization (RHC) in the initial care of critically ill patients (Connors 1996). For the purpose of this paper, we are performing a propensity score analysis to match RHC patients to non-RHC patients using 39 covariates. We will perform the propensity score analysis and optimal matching as well as assess the covariate balance pre and post matching.

COVARIATE EXAMINATION

The study of interest seeks to assess the efficacy of a right heart catheterization (RHC) in the initial care of critically ill patients. This cohort contains 5,735 patients, 2,184 in the treatment group (RHC) and 3,551 in the control group (no RHC). In order to assess an appropriately balanced cohort, we perform a propensity score analysis using 39 covariates, age, gender, years of education, race (Black/white/other), income (less than $11,000/ $11,000-$24,999/$25,000-$49,999/$50,000+), insurance (Medicaid/Medicare/Medicare & Medicaid/Private/Private & Medicare/no insurance), weight (kilograms), primary disease category (acute respiratory failure/multiorgan system failure/congestive heart failure/other), number of comorbidities, do not resuscitate status on day 1, support model estimate of the probability of surviving 2 months, Duke activity status index, APACHE score, Glasgow coma score, mean blood pressure, white blood cell count, heart rate, respiratory rate, temperature, PaO2/FIO2 ratio, albumin, hematocrit, bilirubin, creatinine, sodium, potassium, PaCo2, pH, cancer (none/localized/metastatic), respiratory diagnosis, cardiovascular diagnosis, neurological diagnosis, gastrointestinal diagnosis, renal diagnosis, metabolic diagnosis, hematologic diagnosis, sepsis diagnosis, trauma diagnosis, and orthopedic diagnosis.

In order to assess balance of baseline variables, a common metric is the standardized difference between the treatment and control group, defined as

$$d = \frac{(\bar{X}_{treatment} - \bar{X}_{control})}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}} \times 100\%$$

Where $\bar{X}_{treatment}$ is the mean of the baseline variable in the treatment group and $\bar{X}_{control}$ is the mean of the baseline variable in the control group and s^2 indicates the sample variance. Table 1 displays the baseline characteristics prior to propensity score matching. This table was created using the %PMDIAG macro available at http://www.pharmasug.org/proceedings/2014/SP/PharmaSUG-2014-SP07.pdf (Hulbert 2014).

Variable	No RHC (n=3,551) Mean	RHC (n=2,184) Mean	Difference	Standardized Difference (%)
Age	61.7609	60.7498	1.0111	6.14
Female	0.4610	0.4148	0.0462	9.31
Years of education	11.5690	11.8564	-0.2874	-9.14
Race				
Black	0.1647	0.1534	0.0114	3.10
White	0.7753	0.7816	-0.0063	-1.52
Other	0.0600	0.0650	-0.0050	-2.08
Income				
Less than $11,000	0.5860	0.5243	0.0618	12.45
$11,000-$24,999	0.2008	0.2070	-0.0062	-1.53
$25,000-$49,999	0.1408	0.1799	-0.0391	-10.68
$50,000+	0.0724	0.0888	-0.0165	-6.05
Insurance				
Medicaid	0.1279	0.0884	0.0395	12.74
Medicare	0.2667	0.2340	0.0327	7.55
Medicare & Medicaid	0.0707	0.0563	0.0144	5.89
Private	0.2723	0.3347	-0.0624	-13.60
Private & Medicare	0.2101	0.2244	-0.0143	-3.46
No insurance	0.0524	0.0623	-0.0099	-4.26
Weight (kilograms)	65.0402	72.3602	-7.3199	-25.57
Primary disease category				
Acute Respiratory Failure	0.4452	0.4162	0.0290	5.86
Multiorgan system failure	0.2163	0.3929	-0.1766	-39.09
Congestive heart failure	0.0696	0.0957	-0.0261	-9.50
Other	0.2689	0.0952	0.1737	46.19
Number of comorbidities	1.5207	1.4812	0.0395	3.44
DNR status on day 1	0.1405	0.0710	0.0696	22.76
Support model estimate of the prob. of surviving 2 months	0.2442	0.2020	0.0422	21.29
Duke Activity Status Index	20.3715	20.7008	-0.3293	-6.26
APACHE score	50.9335	60.7390	-9.8055	-50.14
Glasgow Coma Score	22.2532	18.9734	3.2797	10.98
Mean blood pressure	84.8686	68.1978	16.6708	45.51
White blood cell count	15.2635	16.2657	-1.0022	-8.36
Heart rate	112.9	118.9	-6.0551	-14.69
Respiratory rate	28.9781	26.6516	2.3266	16.55
Temperature	37.6329	37.5947	0.0382	2.14
PaO2/FIO2 ratio	240.6	192.4	48.1932	43.32
Albumin	3.1635	2.9776	0.1859	22.99
Hematocrit	32.6997	30.5091	2.1906	26.93
Bilirubin	1.9973	2.7057	-0.7083	-14.46

Variable	No RHC (n=3,551) Mean	RHC (n=2,184) Mean	Difference	Standardized Difference (%)
Creatinine	1.9236	2.4734	-0.5498	-26.96
Sodium	137.0	136.3	0.7043	9.22
Potassium	4.0773	4.0495	0.0277	2.71
PaCo2	39.9526	36.7920	3.1606	24.86
PH	7.3935	7.3802	0.0132	11.98
Cancer				
None	0.7468	0.7908	-0.0439	-10.43
Localized	0.1797	0.1529	0.0267	7.18
Metastatic	0.0735	0.0563	0.0172	6.98
Respiratory Diagnosis	0.4171	0.2894	0.1277	26.95
Cardiovascular Diagnosis	0.2836	0.4231	-0.1395	-29.49
Neurological Diagnosis	0.1619	0.0540	0.1079	35.30
Gastrointestinal Diagnosis	0.1470	0.1923	-0.0453	-12.09
Renal Diagnosis	0.0414	0.0678	-0.0264	-11.63
Metabolic Diagnosis	0.0484	0.0426	0.0059	2.81
Hematologic Diagnosis	0.0673	0.0527	0.0146	6.17
Sepsis Diagnosis	0.1450	0.2363	-0.0912	-23.38
Trauma Diagnosis	0.00507	0.0156	-0.0105	-10.40
Orthopedic Diagnosis	0.000845	0.00183	-0.0010	-2.70

Table 1. Baseline characteristics of 5,735 patients.

As a rule of thumb, standardized differences between the treatment and control groups greater than 10% suggest substantial imbalance (Austin, 2009), however imbalance of any magnitude may be important to consider. Examining Table 1, we see many of the variables have standardized differences of a high magnitude. We can examine the distribution of the covariates by treatment group more closely using UNIVARIATE procedure. For example, the standardized difference for mean blood pressure between the treatment and control group is 45.53%. To examine the overlap in these distributions, we can use the following code:

```
PROC UNIVARIATE DATA=data;
    CLASS treat;
    VAR meanbp;
    HISTOGRAM meanbp/odstitle=”Distribution of Mean Blood Pressure”;
RUN;
```

Alternatively, if we are interested in viewing the histograms overlaid, we can use PROC SGPLOT.

```
PROC SGPLOT DATA=data;
 XAXIS LABEL = "Mean Blood Pressure";
 HISTOGRAM meanbp/GROUP=treat TRANSPARENCY=.5;
 FORMAT treat tgroup.;
RUN;
```

The `XAXIS` statement is used to implement a `LABEL` for the x-axis. The `HISTOGRAM` statement creates histograms of the treatment and control stratified mean blood pressure. The `TRANSPARENCY=.5` option is used to allow for the overlaid histogram to be transparent to allow an easier comparison between groups.

Figure 1 displays the distribution of mean blood pressure by treatment group. From this, it appears that there is overlap in the distributions despite the large standardized difference, so hopefully the propensity score process can take care of this imbalance.

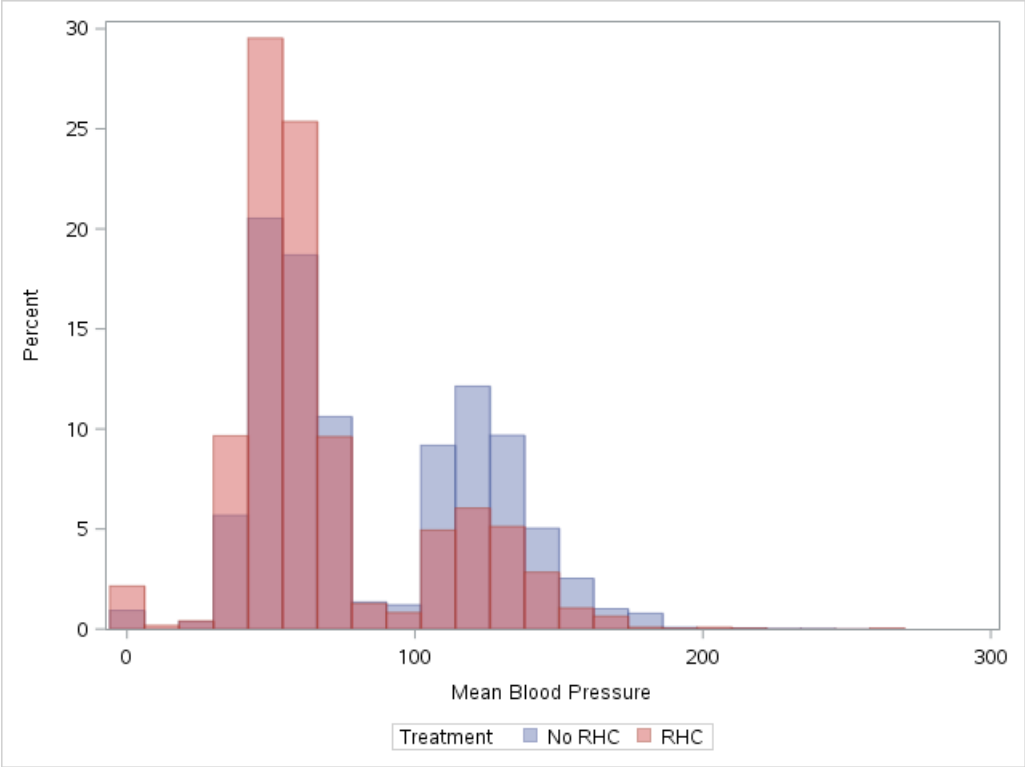

Figure 1. Distribution of mean blood pressure by treatment group.

PROPENSITY SCORE ANALYSIS

PROPENSITY MODEL

To perform the propensity score analysis, we will use logistic regression. All continuous predictors are fit using restricted cubic splines with four knots placed at the 5th, 35th, 65th, and 95th percentiles (Harrell 2015) using the %RCSPLINE macro (available online at http://biostat.mc.vanderbilt.edu/wiki/Main/SasMacros). The following code is used to fit the propensity score model.

```
PROC LOGISTIC DATA=data;
    CLASS catvar1 catvar1 ...;

    MODEL treat(EVENT='1') = var1 var2 ...;

    OUTPUT out = propscores PREDICTED = predpscore XBETA=pscore;
RUN;
```

Here, we enter all categorical predictors in the `CLASS` statement. The `MODEL` statement contains the treatment variable, in this case `treat`, followed by all of the predictors of interest. We use the `OUTPUT` statement to output the propensity scores into a dataset called `propscores`. The variable `predpscore` is on the probability scale, and `pscore` is on the log odds scale. We will perform matching using `pscore`, which gives better differentiation in the tails of the distribution. After initially fitting this model, two subjects had extreme propensity scores, one near 0 and one near 1 on the probability scale. This suggested a potential violation of the positivity assumption, thus these subjects were trimmed (removed from the cohort) and the propensity score model was refit with the remaining 5,733 patients.

DISTRIBUTION OF PROPENSITY SCORES

To examine the distribution of propensity scores, we will overlay histograms using PROC SGPLOT.

```
PROC SGPLOT DATA=propscores;
 XAXIS LABEL = "Log Odds Propensity Score";
 HISTOGRAM pscore/GROUP=treat TRANSPARENCY=.5;
 FORMAT treat tgroup.;
RUN;
```

Examining Figure 2, it appears that there is a region of overlap between the two groups, however there are also observations on either side that are unlikely to find good matches.

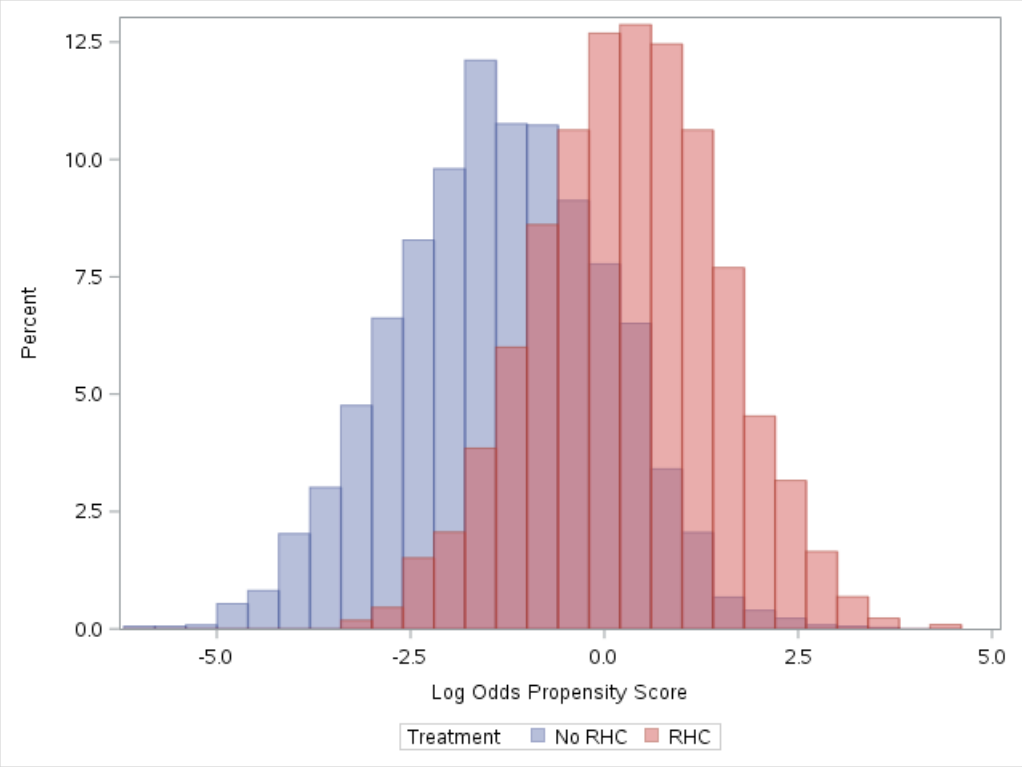

Figure 2. Overlaid histogram of propensity scores for No RHC versus RHC.

MATCHING

We will implement a 1:1 optimal match.

INTEGRATING SAS AND R

In order to perform optimal matching using the nbpMatching package in R, we first need to determine whether you have permission to call R from within SAS by running the following code:

```
PROC OPTIONS OPTION=RLANG;
RUN;
```

If you have permissions, you will see the following in the log, “RLANG Enables SAS to execute R language statements.” If you do not have permissions, you will see “NORLANG”. If you see the latter, save your program and reopen SAS with the RLANG option. This can be done by adding –RLANG to the SAS configuration file, or by launching SAS from the command line in the following manner:

```
“C:\Program Files\SASHome\SASFoundation\9.4\sas.exe” -RLANG
```

Now that we have established that R is implementable, we will begin to use the IML procedure to implement R code. To begin the IML procedure run:

```
PROC IML;
```

Upon completing the code in this section, we will use `QUIT;` to exit this procedure.

There are four important lines of code when implementing R in SAS:

1. `run ExportDataSetToR("sasData", "rData" );`
2. `submit / R;`
3. `endsubmit;`
4. `run ImportDataSetFromR ("rData", "sasData");`

The first line will export a SAS data set, in this case named `sasData` into a data set in R, in this case named `rData`. The second line will indicate the beginning of R code, the third line will indicate the completion of R code, and the fourth line will import an R data set, in this case named `rData`, into a SAS data set named `sasData`.

We first want to pull our dataset of propensity scores, `propscores` into the R environment. To do this we will implement the `ExportDataSetToR` command.

```
run ExportDataSetToR("propscores", "propscores" );
```

Now that we have read our propensity scores into R, we can implement the nbpMatching package to perform optimal matching.

OPTIMAL MATCHING PROCEDURE

Before using the nbpMatching procedure, we need to install this package in R. This can be done within SAS or externally. To install an R package through SAS:

```
submit / R;
   install.packages("nbpMatching")
endsubmit;
```

Once the package is installed, we will implement the matching. The following code implements the matching.

```
submit / R;
   # read library
    library(nbpMatching)

    # select id, treatment as 0/1 indicator, prop score on log odds scale
    d1 <- propscores[,c("ptid","treat","pscore")]

    # calculate caliper
    caliper <- 0.2 * sd(d1[,3])

    # create distance matrix
    d2 <- gendistance(d1,idcol=1,prevent=2)

    # set distances > caliper to infinity
    d3 <- d2$dist
    d3[d3 > caliper] <- Inf
```

```
    # create matches
    d4 <- nonbimatch( distancematrix(d3) )

    # remove mates with Inf distances and the one match to a phantom if N
    # was odd
    d5 <- d4$halves
    d5 <- d5[d5[,5]!=Inf,]
    drop <- c(grep('phantom',d5[,1]),grep('phantom',d5[,3]))
    if(length(drop)>0) d5 <- d5[-drop,]

    # collect IDs and subset to matched cohort
    matchedRows <- c( d5[,2], d5[,4] )
    matched<-data.frame(ptid=propscores[matchedRows,"ptid"])
endsubmit;
run ImportDataSetFromR ("matched", "matched");
QUIT;
```

We first call the `nbpMatching` library using the `library(nbpMatching)` command. We reduce the dataset to include three variables, patient ID (`ptid`), treatment indication (`treat`), and propensity score (`pscore`). We then create a caliper of 0.2 of the pooled standard deviations, as suggested in the literature (Austin 2011, Rosenbaum and Rubin 1985). We create a distance matrix using the `gendistance()` function. The first input is our reduced dataset, the second, `idcol=1`, indicates that our ID variable is in the first column, and the third input, `prevent=2`, indicates that our treatment group is in the second column. The `prevent` option restricts matches between treatment and control groups, rather than allowing any patient to match to any other patient. In order to implement the desired caliper, we set any distance greater than the caliper to infinity. We then create matches using the `nonbimatch(distancematrix())` functions. The `distancematrix()` function reformats the input distance matrix into the format required by the `nonbimatch()` function. We then remove any matches that have an infinite distance, indicating that a match did not exist for these patients. If the number of observations is odd, the matching function will input a phantom match that also needs to be removed. Finally, we will collect the matched IDs from the second and fourth column of the output and put them in a data frame named `matched`. We import this dataset of matched IDs back into SAS using the `ImportDataSetFromR` command.

DISTRIBUTION OF MATCHED PROPENSITY SCORES

Using similar code as implemented above, we examine the distribution of the propensity scores for the matched cohort (Figure 3). We see the overlap in propensity scores is greatly improved as compared to Figure 2.

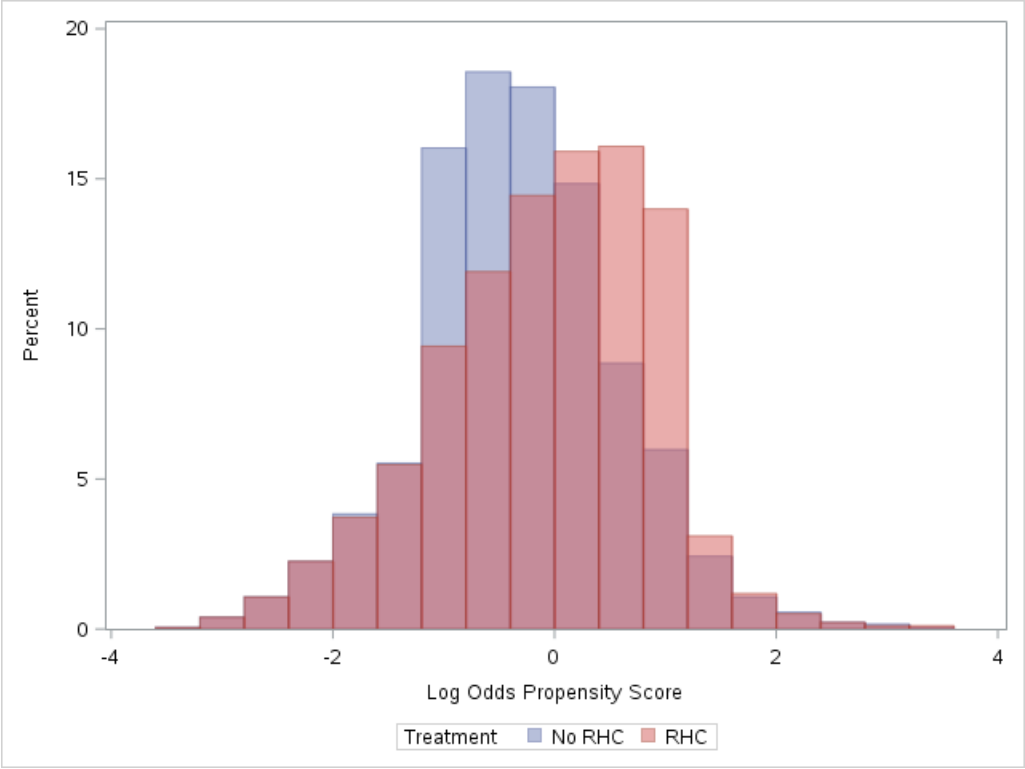

Figure 3. Overlaid histogram of matched cohort propensity scores for No RHC versus RHC.

ASSESSMENT OF MATCH BALANCE

The 1:1 optimal matching resulted in 1,771 matched pairs. To assess match balance, we can again examine standardized differences (Table 2). The standardized differences of the matched cohort are much improved, with no differences greater than 10%, suggesting that balance has been improved.

Variable	No RHC (n=1,771) Mean	RHC (n=1,771) Mean	Difference	Standardized Difference (%)
Age	60.7802	60.5365	0.2437	1.51
Female	0.4210	0.4244	-0.0034	-0.69
Years of education	11.7663	11.8144	-0.0481	-1.53
Race				
Black	0.1603	0.1541	0.0062	1.71
White	0.7743	0.7839	-0.0096	-2.31
Other	0.0655	0.0621	0.0034	1.39
Income				
Less than $11,000	0.5468	0.5356	0.0113	2.26
$11,000-$24,999	0.2133	0.2088	0.0045	1.11
$25,000-$49,999	0.1597	0.1716	-0.0119	-3.19
$50,000+	0.0801	0.0841	-0.0040	-1.44
Insurance				
Medicaid	0.1005	0.0942	0.0062	2.09
Medicare	0.2353	0.2359	-0.0006	-0.13
Medicare & Medicaid	0.0604	0.0587	0.0017	0.72
Private	0.3172	0.3251	-0.0079	-1.69
Private & Medicare	0.2297	0.2252	0.0045	1.08

Variable	No RHC (n=1,771) Mean	RHC (n=1,771) Mean	Difference	Standardized Difference (%)
No insurance	0.0570	0.0609	-0.0040	-1.68
Weight (kilograms)	70.2384	70.8538	-0.6154	-2.27
Primary disease category				
Acute Respiratory Failure	0.4594	0.4351	0.0243	4.88
Multiorgan system failure	0.3087	0.3459	-0.0372	-7.94
Congestive heart failure	0.1072	0.1055	0.0017	0.55
Other	0.1247	0.1134	0.0113	3.48
Number of comorbidities	1.5135	1.4994	0.0141	1.22
DNR status on day 1	0.0886	0.0784	0.0102	3.67
Support model estimate of the prob. of surviving 2 months	0.2230	0.2177	0.0053	2.66
Duke Activity Status Index	20.6665	20.5761	0.0904	1.75
APACHE score	56.5790	58.0350	-1.4560	-7.41
Glasgow Coma Score	18.5322	18.9165	-0.3843	-1.35
Mean blood pressure	73.7813	71.7579	2.0234	5.71
White blood cell count	15.8388	15.8364	0.0023	0.02
Heart rate	116.2	117.1	-0.9944	-2.41
Respiratory rate	28.1558	27.7229	0.4328	3.08
Temperature	37.6079	37.6228	-0.0149	-0.83
PaO2/FIO2 ratio	211.8	206.0	5.7660	5.38
Albumin	3.0458	3.0197	0.0261	3.14
Hematocrit	30.9821	30.8110	0.1711	2.21
Bilirubin	2.5514	2.6159	-0.0645	-1.21
Creatinine	2.2419	2.3126	-0.0706	-3.45
Sodium	136.5	136.5	-0.0119	-0.16
Potassium	4.0435	4.0376	0.0059	0.58
PaCo2	37.7721	37.2361	0.5360	4.71
PH	7.3895	7.3869	0.0026	2.34
Cancer				
None	0.7613	0.7754	-0.0141	-3.34
Localized	0.1721	0.1631	0.0090	2.42
Metastatic	0.0666	0.0615	0.0051	2.07
Respiratory Diagnosis	0.3488	0.3155	0.0333	7.07
Cardiovascular Diagnosis	0.3691	0.3962	-0.0271	-5.57
Neurological Diagnosis	0.0734	0.0638	0.0096	3.80
Gastrointestinal Diagnosis	0.1840	0.1840	0.0000	0.00
Renal Diagnosis	0.0570	0.0626	-0.0056	-2.38
Metabolic Diagnosis	0.0468	0.0429	0.0040	1.91
Hematologic Diagnosis	0.0626	0.0598	0.0028	1.18
Sepsis Diagnosis	0.1992	0.2229	-0.0237	-5.81
Trauma Diagnosis	0.00903	0.0102	-0.0011	-1.16
Orthopedic Diagnosis	0	0.00113	-0.0011	-4.75

Table 2. Baseline characteristics of 3,542 matched patients.

We can also examine the distribution of the covariates in the matched sample compared to the pre-matched sample. After concatenating the matched cohort with the pre-matched cohort and including an indicator for whether a patient is in the matched cohort, we will use PROC SGPANEL to compare the two:

```
PROC SGPANEL DATA=fullandmatch;
PANELBY match;
 COLAXIS LABEL = "Mean Blood Pressure";
 HISTOGRAM meanbp/GROUP=treat TRANSPARENCY=.5;
 FORMAT treat tgroup. match match.;
RUN;
```

Looking at mean blood pressure again, we see that the matching greatly improved the overlap in distributions (Figure 4).

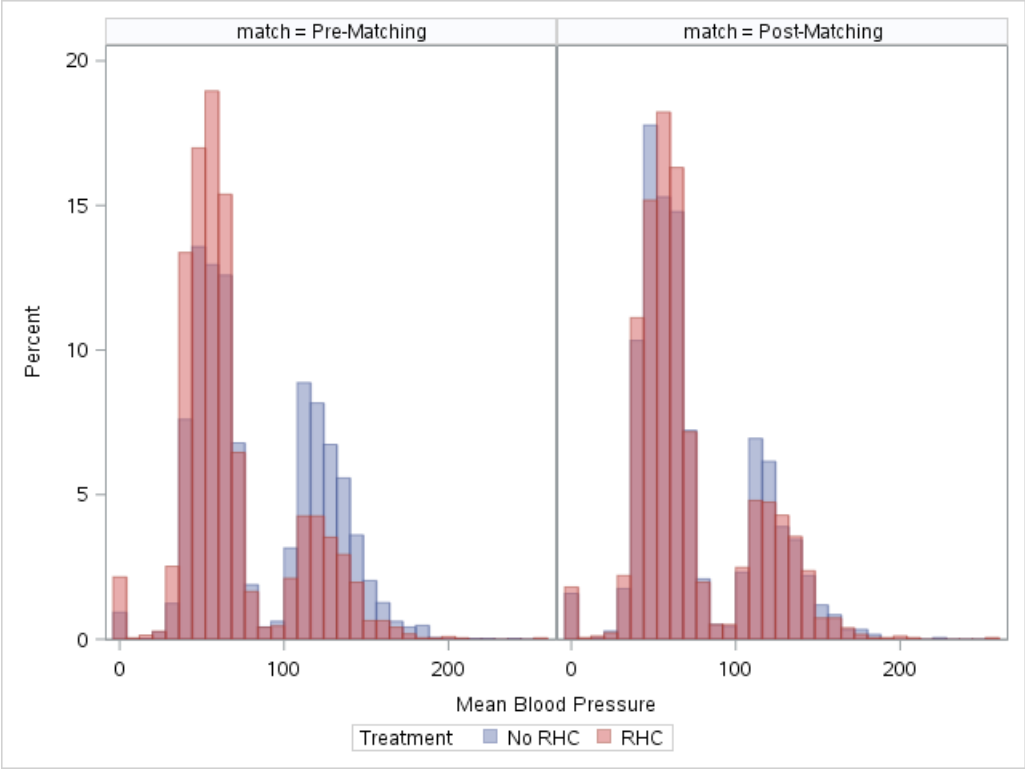

Figure 4. Distribution of mean blood pressure in the pre-matched cohort (n=5,735) compared to the matched cohort (n=3,542).

CONCLUSION

Propensity score matching can be used to reduce bias in observational studies. This paper outlines methods to implement propensity score analyses in SAS® coupled with the implementation of optimal matching using R. The ability to integrate these two languages in a single program streamlines processes and reduces data errors, increasing efficiency and reproducibility.

REFERENCES

Austin, P.C. 2009. “Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples.” *Statistics in Medicine*, 28, 3083-3107.

Austin, P.C. 2011. “Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies.” *Pharmaceutical Statistics*, *10*(2), 150–161.

Beck, C. Lu, B., Greevy R. 2016. nbpMatching: Functions for Optimal Non-Bipartite Matching. R package version 1.5.0. http://CRAN.R-project.org/package=nbpMatching

Connors, A.F., Speroff, T., Dawson, N.V., Thomas, C., Harrell, F.E., Wagner, D., Desbiens, N., Goldman, L., Wu, A.W., Califf, R.M. and Fulkerson, W.J., 1996. “The effectiveness of right heart catheterization in the initial care of critically III patients.” *Jama*, *276*(11), 889-897.

D'Agostino, R.B., Jr., 1998. “Tutorial in Biostatistics: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group", *Statistics in Medicine*, 17, 2265-2281.

Harrell, F. 2015. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Hulbert, E., Brekke, L. 2014. "A SAS® Macro to Evaluate Balance after Propensity Score Matching." *Proceedings of PharmaSUG 2014 Conference*. Available online at: http://www.pharmasug.org/proceedings/2014/SP/PharmaSUG-2014-SP07.pdf.

Lu, B., Greevy, R., Xu, X., Beck, C. 2011. "Optimal Nonbipartite Matching and Its Statistical Applications." *The American Statistician*, 65(1), 21–30.

Rosenbaum, P.R., Rubin, D.B. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity score." *The American Statistician*, 39(1), 33-38.

Rosenbaum, P.R., 1989. "Optimal Matching for Observational Studies" *Journal of the American Statistical Association*, 84(408), 1024-1032.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Lucy D'Agostino McGowan
ld.mcgowan@vanderbilt.edu