



Biostatistics, Epidemiology, and Research Design (BERD) Core

Tabitha Garwe, PhD

Clinical Epidemiology Unit

Propensity Score Analysis in Epidemiologic Research

PART II: INTEGRATION APPROCHES

May 25th, 2022

There are several means of integrating propensity scores into an observational analysis. The three most common techniques are matching, stratification (also called sub-classification) and regression adjustment. Each of these techniques is a way to make an adjustment for covariates prior to (matching and stratification) or while (stratification and regression adjustment) calculating the treatment effect. With all three techniques, the propensity score is calculated the same way, but once it is estimated it is applied differently. Continuing with our prior example of whether an LSI confers a survival advantage, we will explore these three main techniques in integrating the propensity score into an analysis. Use dataset generated from our earlier example (ranks.sas7bdat) and SAS programs provided.

PART I: REGRESSION (COVARIANCE) ADJUSTMENT

1a. We will start by assessing interaction between our treatment variable (LSI) and propensity score variable. We can use the Breslow-Day test or include an interaction term in our multivariable model

Breslow-Day Test for Homogeneity of the Odds Ratios	
Chi-Square	0.6181
DF	3
Pr > ChiSq	0.8923

1b. If no interaction between PS variable and treatment variable, we can proceed with regression adjustment, the propensity score can be included directly as a probability ranging from 0 to 1 or as an ordinal variable, PS quantiles (I prefer quintiles as long as there is adequate overlap across

all quintiles). In this setting, the propensity score serves as a composite confounder, which reduces bias by adjusting for the pattern of observed confounders between the treatment groups when modeling the dependent variable. The regression analysis may include only the treatment variable and the propensity score as covariates, *or additional important covariates*.

Model must be correctly specified, assess all important interactions and other important assumptions e.g. linear relationship between a continuous variable and the outcome of interest

Joint Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
LSI	1	0.0003	0.9856
quintile	1	0.9159	0.3385
DSev1	1	5.9864	0.0144
DSev2	1	3.5355	0.0601
quintile*LSI	1	3.2530	0.0713
DSev1*LSI	1	0.0228	0.8801
DSev2*LSI	1	4.6639	0.0308
DSev1*DSev2	1	0.0300	0.8624

Assuming no interactions and a correctly specified model

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	1.913	0.972	3.765
quintile	1.0000	1.317	0.891	1.945
DSev1	1.0000	0.506	0.420	0.611
DSev2	1.0000	1.049	1.017	1.082

We can also add other important variables

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	1.676	0.842	3.335
quintile	1.0000	1.308	0.882	1.938
DSev1	1.0000	0.642	0.515	0.802
DSev2	1.0000	1.036	1.004	1.070
DSev1F	1.0000	0.757	0.640	0.896

Alternatively, model raw propensity score variable and additional important variables

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	
Intercept	1	1.0604	1.1102	0.9124	0.3395	
LSI	1	0.6732	0.3512	3.6735	0.0553	
pscore	1	-0.8232	1.5560	0.2799	0.5968	
DSev1	1	-0.5060	0.1264	16.0288	<.0001	
DSev2	1	0.0594	0.0237	6.2800	0.0122	
DSev1F	1	-0.2884	0.0877	10.8095	0.0010	

Assumptions – distribution of outcome is binomial; effects are additive and pscore has a linear effect, no highly correlated variables.

PART II: STRATIFICATION

2a. If adequate overlap across and homogeneity of treatment effect across quintiles as well as, no additional variables to adjust for, then simply calculate a pooled effect estimate; see Part 1a above

Breslow-Day Test for Homogeneity of the Odds Ratios	
Chi-Square	0.6181
DF	3
Pr > ChiSq	0.8923

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel-Haenszel	1.9857	1.0730	3.6749
	Logit **	1.9711	1.0556	3.6804
Relative Risk (Column 1)	Mantel-Haenszel	1.0244	1.0039	1.0454
	Logit	1.0086	0.9923	1.0251
Relative Risk (Column 2)	Mantel-Haenszel	0.5323	0.2982	0.9502
	Logit **	0.5342	0.2984	0.9564

2b. Stratified regression – need to adjust for additional variables and using stratifying variable ‘quintile’

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	1.953	0.998	3.823
DSev1	1.0000	0.490	0.403	0.596
DSev2	1.0000	1.054	1.021	1.089

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	1.727	0.874	3.414
DSev1	1.0000	0.620	0.493	0.780
DSev2	1.0000	1.042	1.008	1.077
DSev1F	1.0000	0.762	0.645	0.900

One can also run a weighted regression – SEE PROC PSMATCH SAMPLE CODE (Sample programs at the bottom of your provided SAS program). In the outcome analysis, you can use the weighted average of the stratum-specific treatment estimates to estimate the treatment effect. You can estimate the ATT if you weight by the stratum-specific number of treated units, and you can estimate the ATE if you weight by the stratum-specific number of units (treated and control units combined) (Stuart 2010, p. 13; Guo and Fraser 2015, pp. 76–77).

PART III: PROPENSITY SCORE MATCHING

Propensity score matching is generally viewed as the most statistically efficient method of incorporating propensity scores, *but requires a large sample size and eliminates unmatched subjects*. Matched sets of treated and untreated subjects share a similar value of the propensity score (Rosenbaum & Rubin, 1983a, 1985) and allows one to **estimate the ATT** (Imbens, 2004).

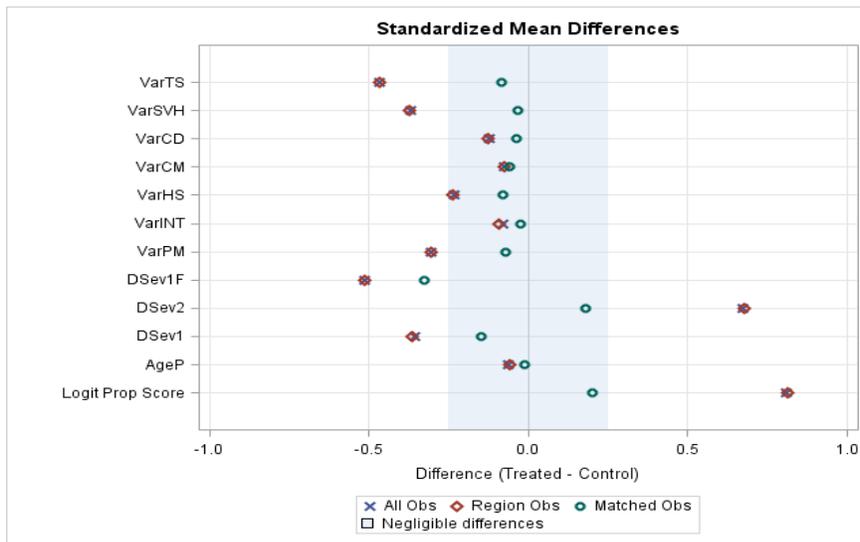
We will explore a few ways to conduct propensity score matching.

3a. Mahalanobis Distance Matching within calipers defined by logit of the propensity score

- Greedy matching, lps caliper = 0.5 (too wide, for demo only)

WARNING: Some treated units do not have matched controls because there are not enough available controls for these treated units.

- NOTE: The data set WORK.MATCHED_MH has 1068 observations and 29 variables.



Evaluation of standardized differences suggest some residual imbalance in variables DSev1F and to some extent DSev2 as well as our propensity score (logit of propensity score); we would consider refitting the propensity score model if possible or additional adjustment in the outcome model

Sample of matched observations

Obs	LSI	DSev1	DSev2	DSev1F	AgeP	_PS_	_Lps	_MatchID
1	0	4.094	35	1.465	29	0.92636	2.53208	1
2	1	2.198	59	1.465	19	0.92408	2.49909	1
3	0	5.030	43	2.930	18	0.88592	2.04972	2
4	1	5.967	50	2.930	27	0.88837	2.07420	2
5	1	4.740	34	2.930	21	0.88374	2.02833	3

Obs	LSI	DSev1	DSev2	DSev1F	AgeP	_PS_	_Lps	_MatchID
6	0	3.221	43	2.930	23	0.83417	1.61545	3
7	0	5.967	34	4.804	63	0.84367	1.68579	4
8	1	5.967	42	2.930	58	0.88367	2.02766	4
9	1	7.550	57	2.930	19	0.87887	1.98179	5
10	0	6.904	29	4.804	21	0.83137	1.59535	5

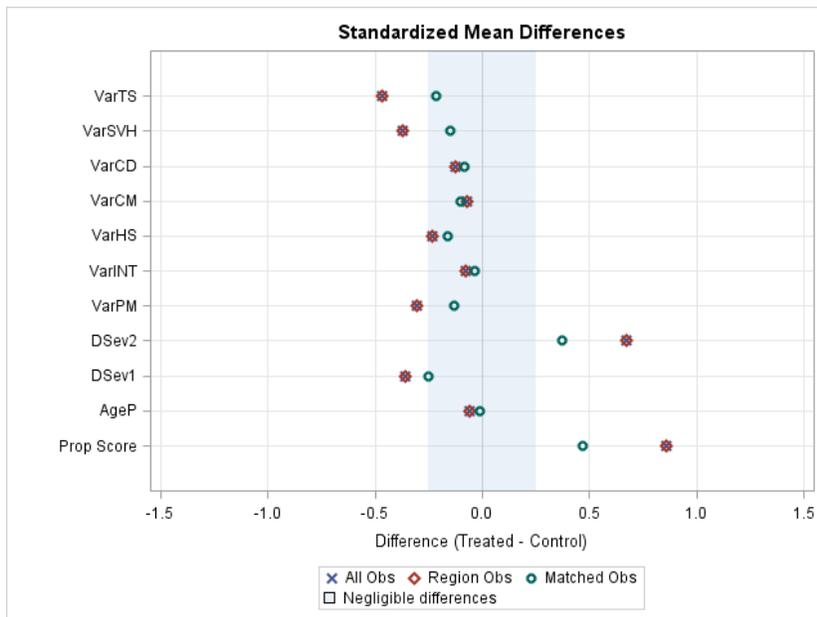
Table of LSI by Died			
LSI	Died		
Frequency Percent Row Pct Col Pct	0	1	Total
0	513 48.86 97.71 50.89	12 1.14 2.29 28.57	525 50.00
1	495 47.14 94.29 49.11	30 2.86 5.71 71.43	525 50.00
Total	1008 96.00	42 4.00	1050 100.00

Conditional Logistic Regression Odds Ratio Estimates and Wald Confidence Intervals

Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	2.060	0.735	5.773
DSev2	1.0000	1.044	0.928	1.175
DSev1F	1.0000	0.232	0.044	1.241

3b. Optimal Fixed Ratio Matching with logit of propensity score - **no defined caliper**

This time **ALL** of our treated units were matched but...we still have residual imbalances (DSev1 DSev2, propensity score)



Sample of matched observations

Obs	LSI	DSev1	DSev2	DSev1F	AgeP	_PS_	_Lps	_MatchID
1	1	7.841	4	7.841	33	0.12135	-1.97968	1
2	0	7.841	9	7.841	43	0.12731	-1.92498	1
3	1	7.841	5	7.841	58	0.13195	-1.88381	2
4	0	7.841	9	7.841	61	0.12864	-1.91308	2
5	0	7.841	1	7.841	57	0.13629	-1.84646	3
6	1	7.841	2	7.841	58	0.14253	-1.79441	3
7	1	7.841	4	7.841	66	0.14271	-1.79297	4
8	0	7.841	1	7.841	94	0.14252	-1.79453	4
9	1	7.841	1	7.841	49	0.14426	-1.78033	5
10	0	7.841	1	7.841	50	0.14392	-1.78310	5

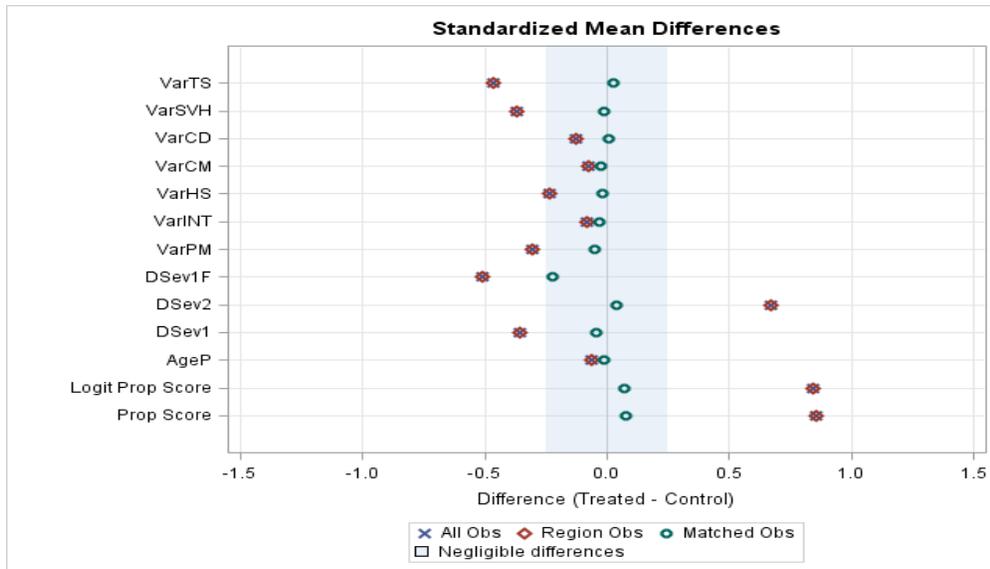
3b. Greedy Matching within calipers defined by logit of propensity score - *lps caliper=0.25*

WARNING: **Some treated units do not have matched controls** because there are not enough available

controls for these treated units.

NOTE: The data set WORK.MATCHGRD has 970 observations and 29 variables. ONLY 37 OF 61 DEATHS INCLUDED

Table of LSI by Died			
LSI	Died		
	0	1	Total
Frequency			
Percent			
Row Pct			
Col Pct			
0	472	13	485
	48.66	1.34	50.00
	97.32	2.68	
	50.59	35.14	
1	461	24	485
	47.53	2.47	50.00
	95.05	4.95	
	49.41	64.86	
Total	933	37	970
	96.19	3.81	100.00



Almost all variables are reasonably balanced, even for DSev1F (although on edge of our acceptable limit – so we will still additionally adjust for it in our outcome model)

Conditional Logistic Regression Odds Ratio Estimates and Wald Confidence Intervals

Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	2.696	0.729	9.968
DSev2	1.0000	1.007	0.917	1.107
DSev1F	1.0000	0.448	0.251	0.798

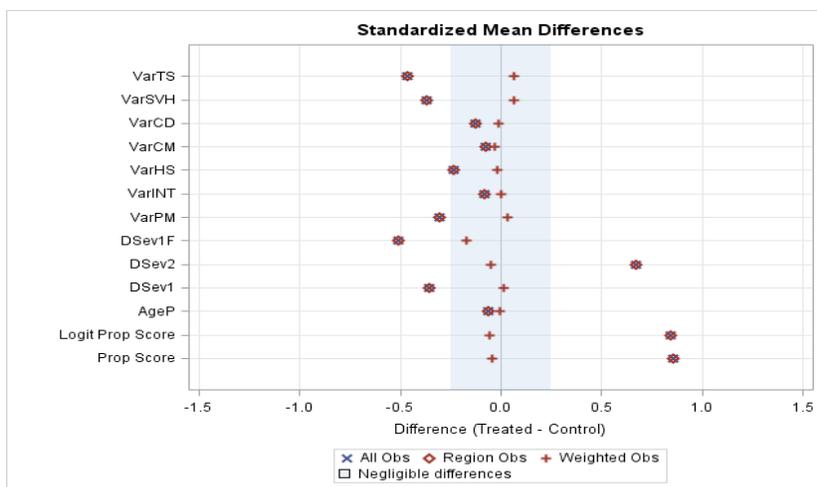
For demo only – a bit difficult trusting these results with reduced sample size that excludes almost half of the deaths (37 of 61 deaths included in model)

/***/TIME-PERMITTING***/

PART IV: PROPENSITY SCORE WEIGHTING

Inverse probability of treatment weighting (IPTW) using the propensity score uses weights based on the propensity score to create a synthetic sample in which the distribution of measured baseline covariates is independent of treatment assignment. The use of IPTW is similar to the use of survey sampling weights that are used to weight survey samples so that they are representative of specific populations (Morgan & Todd, 2008). A subject's weight is *equal to the inverse of the probability of receiving the treatment that the subject actually received*. Inverse probability of treatment weighting was first proposed by Rosenbaum (1987a) as a form of model-based direct standardization. Regression models can be weighted by the inverse probability of treatment to estimate causal effects of treatments. In this context, IPTW is part of a larger family of causal methods known as marginal structural model (Hernan, Brumback, & Robins, 2000, 2002). Variance estimation must account for the weighted nature of the synthetic sample, with robust variance estimation commonly used.

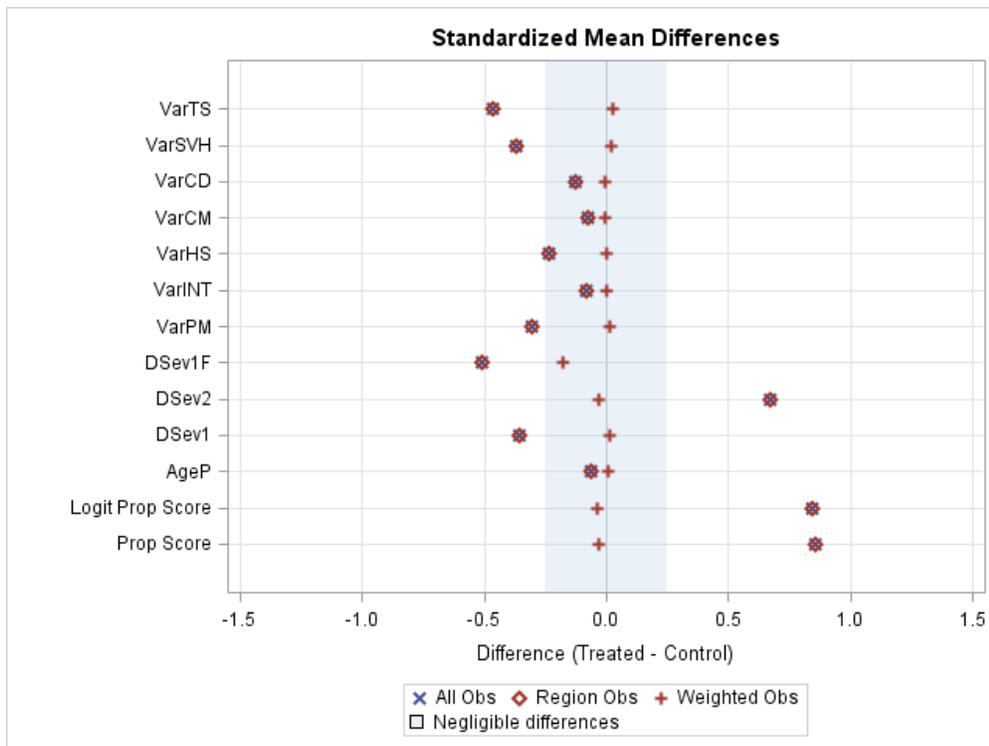
4a. ATT-weighted logistic regression (Proc surveylogistic)



Odds Ratio Estimates and t Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	1.053	0.495	2.239
DSev1F	1.0000	0.501	0.426	0.589

NOTE: The degrees of freedom in computing the confidence limits is 1633.

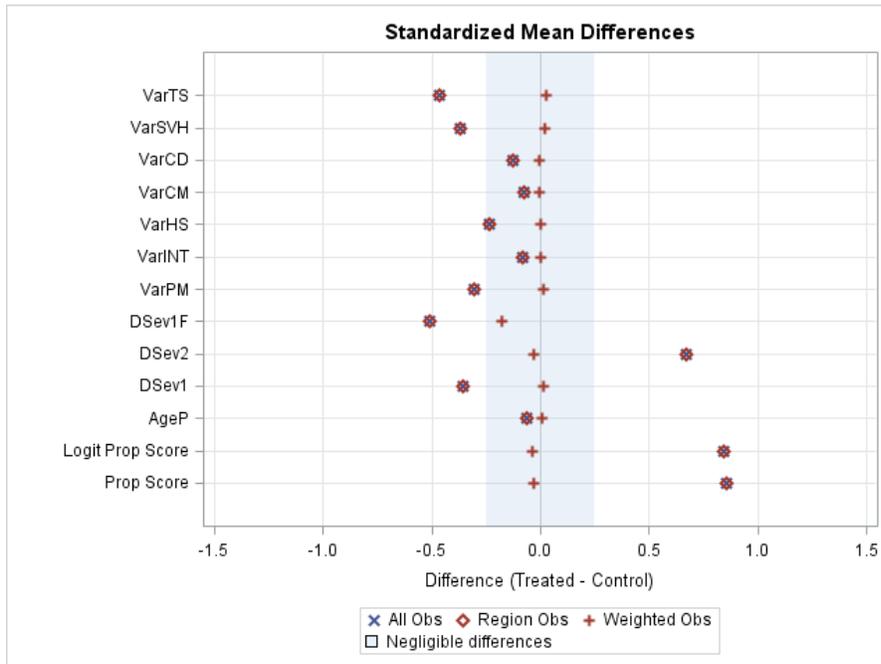
4b. ATE-weighted logistic regression (Proc surveylogistic)



Odds Ratio Estimates and t Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	0.976	0.484	1.965
DSev1F	1.0000	0.493	0.426	0.570

NOTE: The degrees of freedom in computing the confidence limits is 1633.

4c. Stabilized ATE-weighted logistic regression (Proc surveylogistic)



Odds Ratio Estimates and t Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
LSI 1 vs 0	1.0000	0.976	0.484	1.965
DSev1F	1.0000	0.493	0.426	0.570

NOTE: The degrees of freedom in computing the confidence limits is 1633.

References

1. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med.* 1998 Oct 15;17(19):2265-81.
2. Newgard CD, Hedges JR, Arthur M, Mullins RJ. Advanced statistics: the propensity score--a method for estimating treatment effect in observational research. *Acad Emerg Med.* 2004 Sep;11(9):953-61.
3. Austin P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate behavioral research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
4. 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(25), 3083–3107.
5. <https://support.sas.com/documentation/onlinedoc/stat/142/psmatch.pdf>