Paper AS04

How to limit bias in observation studies analysis, Propensity score matching versus Logistic regression

DEBRUS Roxane, Terumo N.V., Leuven, Belgium

ABSTRACT

Because not every scientific question can be answered with randomized controlled trials, research methods that minimize bias in observational studies are required. In the statistical analysis of observational data, propensity score matching (PSM) is a statistical matching technique that attempts to estimate the effect of a treatment or other intervention by accounting for the covariates that predict receiving it. Another statistical method often used in observational data is the multivariate logistic regression (LR) that aims to control for imbalances between groups in order to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. Both strategies, PSM and LR, have their advantages and limitations but complement each other to provide the whole picture. In this article, we, with the aid of a real study example, illustrate different methods to analyze data with selection bias and clustering and with a dichotomous outcome.

INTRODUCTION

Everybody knows that randomized clinical trials have the highest level of scientific evidence and should be performed in clinical research to assess the effect of a treatment between two groups. It is the randomization which assigns the treatment group and therefore ensures a similar distribution of covariables between the groups. However, RCTs are not always possible and representative of reality. To that effect are observational studies, in which subjects are not randomly assigned so covariates are not similarly distributed in both groups. Therefore, methods to minimize bias are required to remove the confounding effect when estimating the effect of the treatment.

In order to do so, there are different methods, but the most common are:

- The **Logistic Regression** (LR) that estimates the treatment effect after adjusting for differences in the baseline covariates
- The **Propensity Score Matching analysis** (PSM) defined as the probability to be assigned to a treatment depending on a set of observed baseline covariates.

When asking around me to previous colleagues or alumni, only a few people knew about PSM and almost no one knew how to execute it. Is it a prehistoric outdated method or a modern analysis method? How come that it is so little known and mastered? It is a bit of both.

The analysis technique had been introduced in 1983 by Paul Rosenbaum in the US. The number of publications published and referenced on Pubmed, suddenly raises in 2005 and from that point increases almost exponentially over the years and reaches 4000 in 2019. In August 2020, there were around 3200 articles published in 2020, and in October 2020, the number reached 3866! We can expect this year to be a new record!

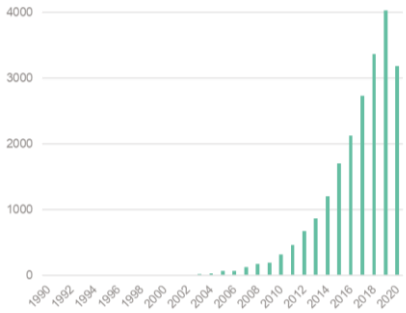


Figure 1 – *Number of publications on Pubmed every year related to PSM.*

If you are wondering why the number of publications using PSM has doubled in the 3 last years, well there are multiple reasons. First, the way of collecting data has been changing over the last 5 years. It is now easier to collect and store big volumes of data from all around the world, such as for registries, for which the Real-World Evidence (RWE) is growing in credibility. Secondly, authorities are demanding post marketing surveillance studies.

As PSM can be performed with many different software, it is not limited to pharmaceutical research but it is also expanding to other analytic department such as academic research:

- **SAS**: by using PSMatch procedure, and macro OneToManyMTCH
- **R**: by using the MatchIt package
- **STATA**: by using the user-written psmatch2 or teffects psmatch built in command (available after version 13)
- **SPSS**: Dialog bow for PSM available from the IBM SPSS Statistics Menu

METHODOLOGY

There are 2 major parts in the process, the first one is the generation of the propensity score, and the second one is how to use this score and integrate it in the analysis.

To generate the scores, first we need to check if there are some imbalances between the 2 treatment groups. Then these imbalances will be used to predict the use of a specific treatment with a logistic regression model. Once the model can be considered as final and adequate, Propensity score can be output. There are many ways to integrate this propensity score in the analysis, and many options to tailor your needs. You can for example adjust, stratify or match subjects. Once this PS have been integrated you need to verify that the imbalances between the groups have been reduced, and if so, you can finally estimate the treatment effect.

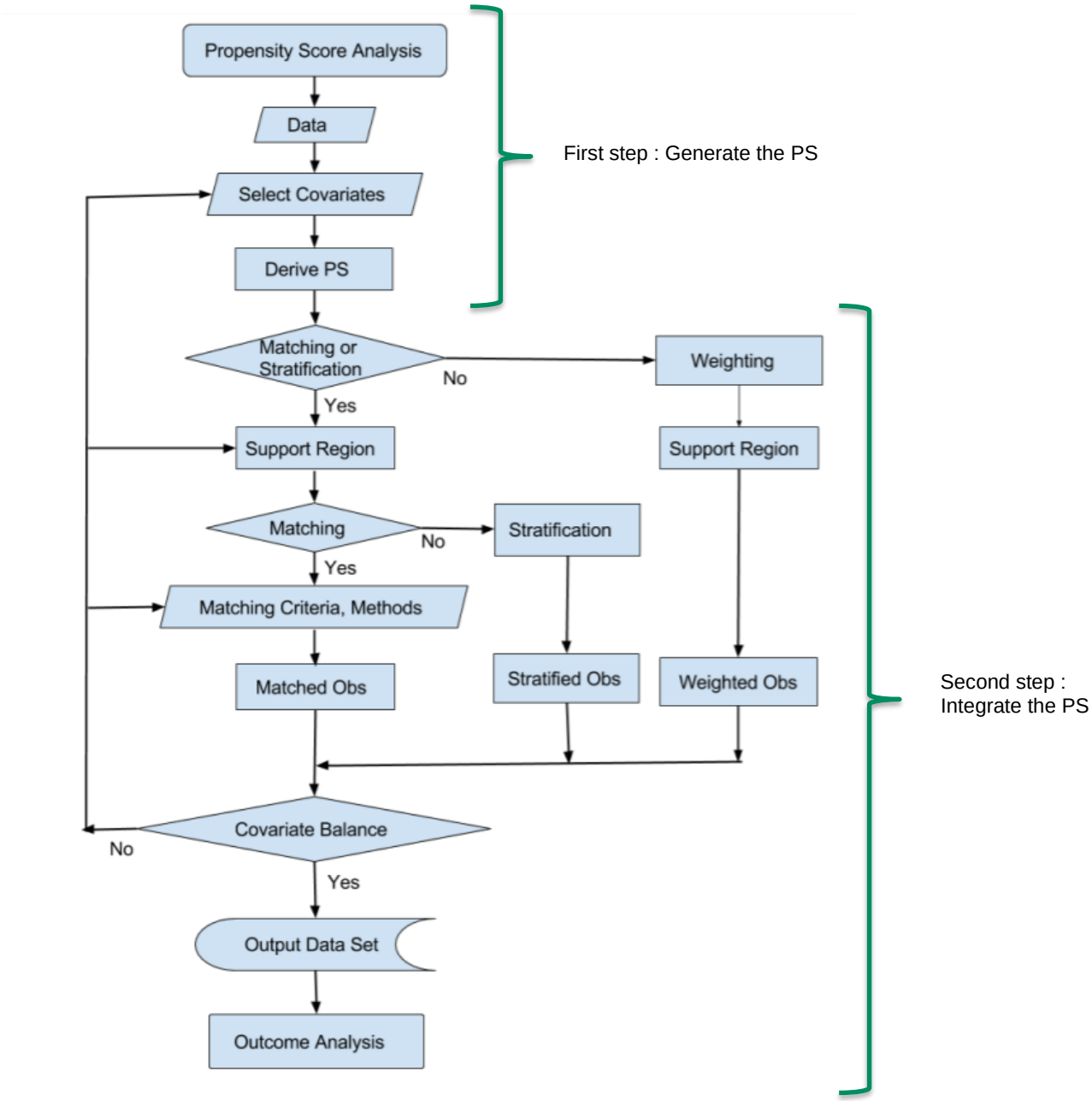
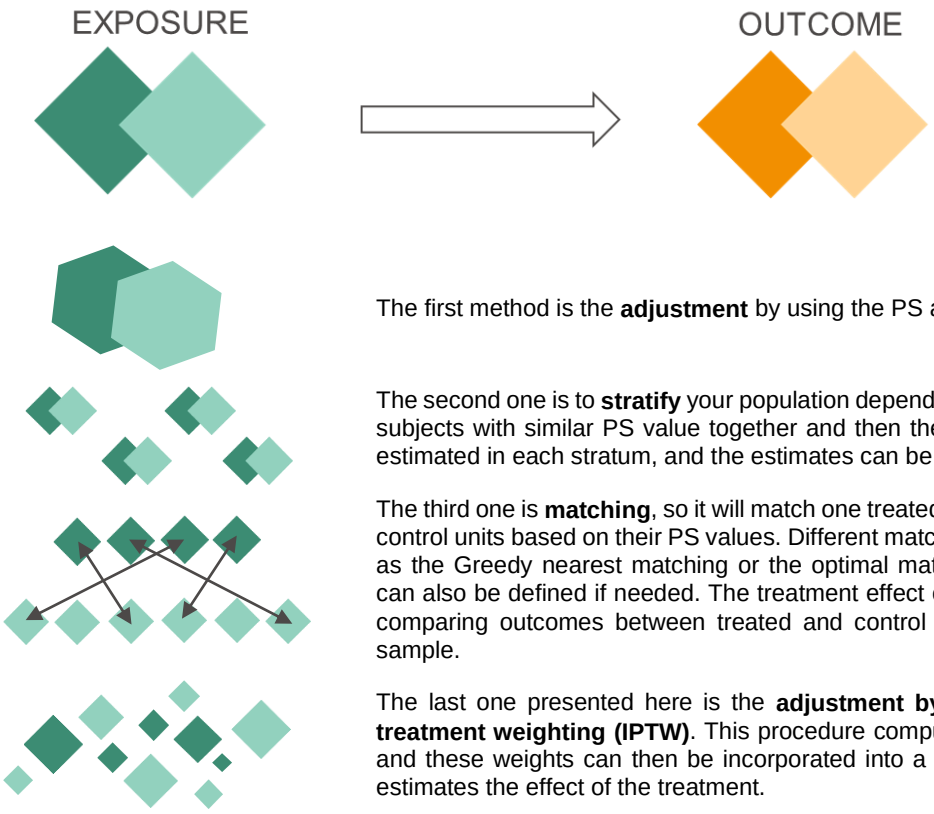


Figure 2 – *Global process of Propensity Score Analysis*

There are 4 major methods to integrate the Propensity Score (PS) in the outcome analysis:



The first method is the **adjustment** by using the PS as a regression covariate.

The second one is to **stratify** your population depending on the PS value to put subjects with similar PS value together and then the treatment effect can be estimated in each stratum, and the estimates can be combined across strata.

The third one is **matching**, so it will match one treated subject with one or more control units based on their PS values. Different matching methods exists such as the Greedy nearest matching or the optimal matching. Additional options can also be defined if needed. The treatment effect can then be estimated by comparing outcomes between treated and control subjects in the matched sample.

The last one presented here is the **adjustment by inverse probability of treatment weighting (IPTW)**. This procedure computes weights from the PS and these weights can then be incorporated into a subsequent analysis that estimates the effect of the treatment.

Figure 3 – R*epresentation of the Effect of Integration of Propensity Score on the exposure*

EXAMPLE ON A TERUMO'S STUDY

Terumo Corporation was founded in 1921 in Japan. Since then Terumo developed more than 100 different medical devices in multiple fields. Currently, Terumo's clinical research is mostly in the fields of Interventional Cardiology, Interventional Oncology, and Peripheral Interventions.

In 2019, Terumo closed its biggest study ever, the e-Ultimaster Trial, which is one of the largest prospective worldwide registries in its field, which enrolled up to 37 000 patients. The device under investigation was a Drug Eluting Coronary Stent. It was an observational study, with a single arm, open label and 5 majors timepoint including baseline, procedure and 1 Year Follow-Up. The primary endpoint was to validate Efficacy and Safety based on a composite endpoint of different serious adverse event up to 1 year after the procedure. Even though the number of subjects enrolled is huge, the study has many limitations as it is observational: all kind of subjects were allowed to be included and patients were treated as per hospital standard of care or Principal Investigator's preferences.

Thanks to the high number of subjects followed up, many sub-analyses were performed. One of them was to investigate the impact of intravascular imaging on the occurrence of target lesion failure up to one year after the procedure. To place a stent in a coronary artery, a puncture is made in the radial artery which is available at the level of the wrist, or in the groin.

Interventional devices are introduced in this hole to reach the target lesion by following the artery up to the heart. Coronary angiography is the standardized imaging technique used during a Percutaneous Coronary Intervention (PCI) to ensure the stent is placed at the right location and to see that the blood flow is back to normal after stenting. These images are obtained from outside of the body (such as an echo doppler for example).

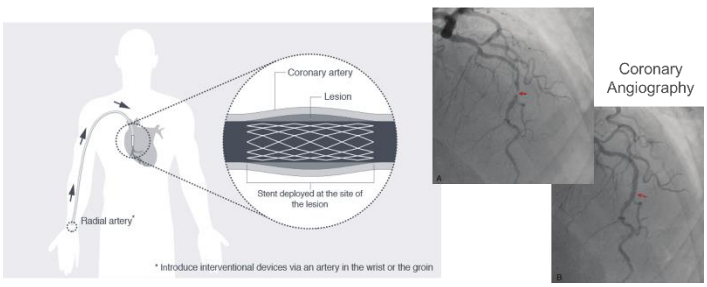


Figure 4 – Puncture site for PCI and Coronary Angiography Imaging

Another imaging technique is the intravascular imaging, which gives direct imaging from the inside of the vessel and can optimize the procedure as it gives valuable additional information to better understand the clinical parameters of the vessel or the lesion. In the eUltimaster study, this type of imaging has been used in 5% of the subjects.

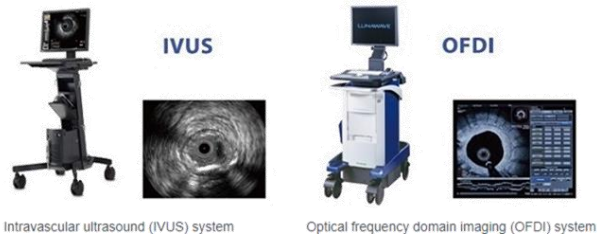


Figure 5 – Intravascular machines and imaging

As this imaging system optimizes the procedure, we expected that subjects for whom it was used to have less post-procedure complications and a better outcome up to 1 year after the procedure.

- X is the exposure, the uses of intravascular imaging: Y or N
- Y is the outcome is the occurrence of target lesion failure up to 1 year after the procedure : Y or N

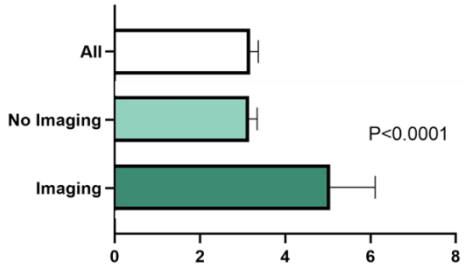


Figure 6 – **TLF Occurence Rate at 1Y**

In the total population, the occurrence rate of Target Lesion Failure (TLF) is equal to 3.2% but once we split in 2 groups depending on the exposure to intravascular imaging, we were quite surprised to see that the rate of TLF up to 1 year was actually higher in the group of subjects where this additional imaging was used compared to the group where it was not used, while we expected the opposite. This difference is statistically significant (p<0.0001).

One of the possible explanations of this finding is that intravascular imaging was probably used to treat more complex cases, and complexity can be defined by subject (such as age, obesity, …) or lesion characteristics (target artery, lesion at bifurcation site, calcified vessels, tortuosity of the vessel, …).

⇨ In order to adjust, we will define U as a set of covariates:

- X is the exposure, the uses of intravascular imaging: Y or N
- Y is the outcome is the occurrence of target lesion failure up to 1 year after the procedure : Y or N
- U is the set of covariates

CHECK FOR IMBALANCES BETWEEN CONTROL AND TREATMENT GROUP

The first step is to check for imbalances between control and treated group. In our case, we identified 10 baseline characteristics and 5 lesion characteristics that were statistically significantly different in the 2 groups (imaging vs no imaging).

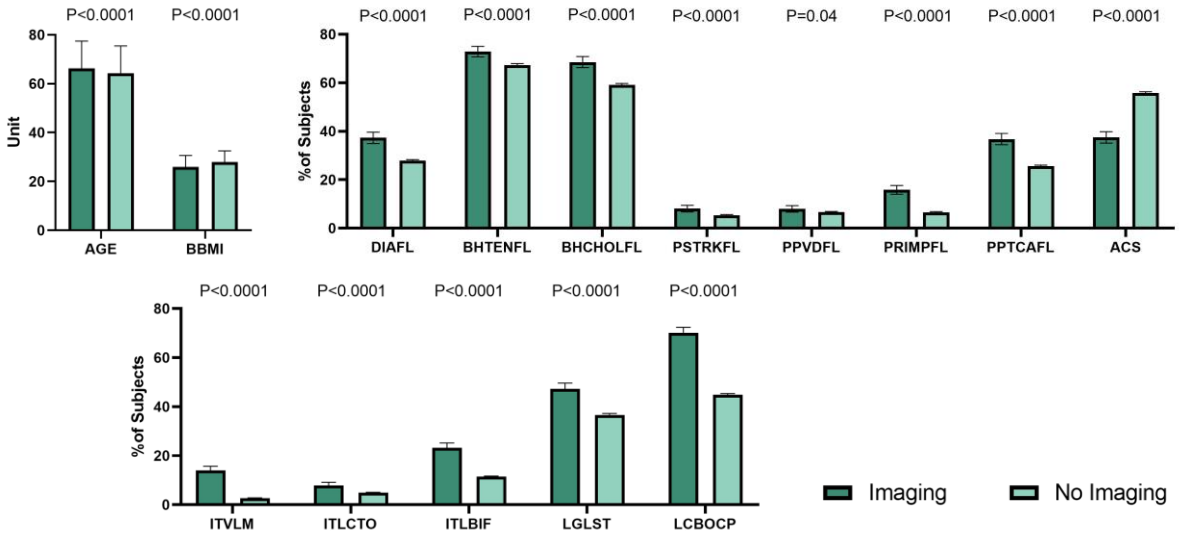


Figure 7 – Imbalances between the 2 exposure groups

Variable	Effect	Number of Observations Read	Number of Observations Used	% of Observations Used	Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio	Pr > Chi-Square
AGE	AGE	35389	35389	100.0%	1.017	1.012	1.021	<.0001
BBMI	BBMI	35389	28446	80.4%	0.898	0.886	0.910	<.0001
DIAFL	DIAFL Y vs N	35389	34780	98.3%	1.539	1.388	1.706	<.0001
BHTENFL	BHTENFL Y vs N	35389	31990	90.4%	1.302	1.163	1.457	<.0001
BHCHOLFL	BHCHOLFL Y vs N	35389	30843	87.2%	1.503	1.349	1.675	<.0001
PSTRKFL	PSTRKFL Y vs N	35389	32847	92.8%	1.552	1.281	1.879	<.0001
PPVDFL	PPVDFL Y vs N	35389	32200	91.0%	1.222	1.014	1.473	0.0356
PRIMPFL	PRIMPFL Y vs N	35389	34628	97.8%	2.659	2.312	3.058	<.0001
PPTCAFL	PPTCAFL Y vs N	35389	32965	93.2%	1.690	1.523	1.876	<.0001
ACS	ACS Y vs N	35389	35369	99.9%	0.474	0.428	0.525	<.0001
ITVLM	ITVLM Y vs N	35389	35389	100.0%	6.027	5.165	7.034	<.0001
ITLCTO	ITLCTO Y vs N	35389	35389	100.0%	1.664	1.381	2.005	<.0001
ITLBIF	ITLBIF Y vs N	35389	35389	100.0%	2.348	2.084	2.645	<.0001
LGLST	LGLST Y vs N	35389	35343	99.9%	1.609	1.457	1.777	<.0001
LCBOCP	LCBOCP Y vs N	35389	35389	100.0%	2.893	2.598	3.223	<.0001

Figure 8 – Detailed output for imbalances between the 2 exposure groups

In this observational study, risk-based monitoring had been implemented, so most of the data was not source-verified. For the data itself, not every variable or question in the system was mandatory for completion. The handling of missing data is an important step and can be solved in different ways. For this example, we imputed missing values in order to keep all subjects in the analysis.

LOGISTIC REGRESSION MODEL

What would have been the analysis results if we would have tried to predict the occurrence of TLF up to 1 year based on the use of intravascular imaging when adjusting for the 15 imbalanced baseline and lesion characteristics? After adjustment for these 15 predictors, the impact of additional intravascular imaging on TLF occurrence up to 1 year is not statistically significant anymore (p=0.43).

Number of Observations Read	35389
Number of Observations Used	35389

TLF1Y	Y	1135
	N	34254

Effect	DF	Wald Chi-Square	Pr > Chi-Square	Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio
Intercept	1	292.8289	<.0001			
IMAGE Y vs N	1	0.6185	0.4316	1.103	0.864	1.410
AGE	1	45.6225	<.0001	1.020	1.014	1.026
BBMI	1	0.8080	0.3687	0.993	0.979	1.008
DIAFL Y vs N	1	27.1735	<.0001	1.410	1.239	1.605
BHTENFL Y vs N	1	1.2212	0.2691	0.925	0.806	1.062
BHCHOLFL Y vs N	1	0.2275	0.6334	0.969	0.853	1.102
PSTRKFL Y vs N	1	5.3868	0.0203	1.295	1.041	1.612
PPVDFL Y vs N	1	16.0479	<.0001	1.492	1.227	1.815
PRIMPFL Y vs N	1	50.4779	<.0001	1.889	1.585	2.251
PPTCAFL Y vs N	1	35.9979	<.0001	1.495	1.311	1.705
ACS Y vs N	1	24.9200	<.0001	1.376	1.214	1.560
ITVLM Y vs N	1	48.3981	<.0001	2.258	1.795	2.840
ITLCTO Y vs N	1	0.2018	0.6533	1.065	0.809	1.401
ITLBIF Y vs N	1	12.4154	0.0004	1.347	1.141	1.589
LGLST Y vs N	1	0.0056	0.9402	1.005	0.886	1.140
LCBOCP Y vs N	1	16.0442	<.0001	1.299	1.143	1.476

Figure 9 – Detailed output for Logistic Regression Model

COMPUTE THE PROPENSITY SCORES

ESTIMATE THE PROPENSITY MODEL

Let's go a bit further and build another model, not to predict the outcome anymore, but to predict the exposure based on these 15 imbalanced baseline and lesion characteristics with a stepwise selection (Method : include all predictors, Stepwise Selection, SLE=0.05, SLS=0.05).

```
proc logistic data=psm.analysis_image_imputed ;
class           IMAGE (ref="N") DIAFL (ref='N') BHTENFL (ref='N')
        BHCHOLFL (ref='N') PSTRKFL (ref='N') PPVDFL (ref='N')
        PRIMPFL (ref='N') PPTCAFL (ref='N') ACS (ref='N')
        ITVLM (ref='N') ITLCTO (ref='N') ITLBIF (ref='N')
        LGLST (ref='N') LCBOCP (ref='N')
        /param = ref;
model   /*X*/ IMAGE =
        /*N*/   AGE BBMI
        /*C*/   DIAFL BHTENFL BHCHOLFL PSTRKFL PPVDFL PRIMPFL
                PPTCAFL ACS ITVLM ITLCTO ITLBIF LGLST LCBOCP
        / selection=stepwise sle=0.05 sls=0.05 ;
Output  out=PS predicted=PRED;
run;
```

In the model, 12 of the 15 variables are selected, and the PS can be output by adding an output statement into the proc logistic procedure.

Number of Observations Read	35389
Number of Observations Used	35389

IMAGE	Y	1642
	N	33747

Step	Effect Entered	DF	Number of Effects In Model	Score Chi-Square	Pr > Chi-Square	Effect	Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio
1	ITVLM	1	1	662.3452	<.0001	ITVLM Y vs N	3.849	3.237	4.577
2	LCBOCP	1	2	333.1108	<.0001	LCBOCP Y vs N	2.232	1.989	2.505
3	BBMI	1	3	272.6651	<.0001	BBMI	0.876	0.864	0.889
4	ACS	1	4	195.5462	<.0001	ACS Y vs N	0.549	0.493	0.611
5	PRIMPFL	1	5	124.5823	<.0001	PRIMPFL Y vs N	1.945	1.671	2.265
6	BHCHOLFL	1	6	102.3030	<.0001	BHCHOLFL Y vs N	1.591	1.421	1.782
7	DIAFL	1	7	48.6366	<.0001	DIAFL Y vs N	1.441	1.290	1.609
8	PPTCAFL	1	8	22.4584	<.0001	PPTCAFL Y vs N	1.300	1.163	1.452
9	LGLST	1	9	19.2665	<.0001	LGLST Y vs N	1.260	1.134	1.400
10	ITLBIF	1	10	17.5167	<.0001	ITLBIF Y vs N	1.333	1.165	1.526
11	BHTENFL	1	11	7.7020	0.0055	BHTENFL Y vs N	1.190	1.056	1.342
12	PPVDFL	1	12	5.5940	0.0180	PPVDFL Y vs N	0.787	0.645	0.960

Figure 10 – Detailed output for Propensity Model

PREDICT INDIVIDUAL PROPENSITY SCORE

By adding the output statement in the propensity model, you can output the propensity score (PRED) in an output dataset (WORK.PS).

The propensity score represents a probability to get a specific exposure and therefore its value is always contained between 0 and 1. In the graph, you can see how these scores are distributed in the 2 treatment groups.

In each overlap, a comparison can be made.

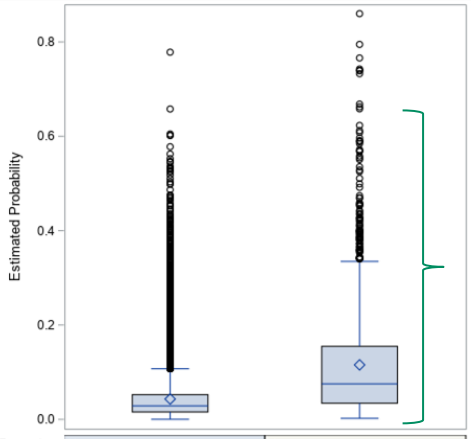


Figure 11 – Distribution of Propensity Score by Exposure

INTEGRATE THE PROPENSITY SCORE IN THE ANALYSIS

Now that we have a PS populated for each subject, how to integrate it in the analysis?

METHOD 1: LOGISTIC REGRESSION WITH PS ADJUSTMENT

The first option was to adjust the logistic regression by using the PS as a regression covariate. So, in the model statement, you will have the outcome (Y) = the exposure (X) and the PS (U).

```
proc logistic data=PS ;
class /*Y*/ TLF1Y (ref='N') IMAGE (ref='N') /param = ref;
model /*Y*/ TLF1Y = IMAGE PRED /*PS - X on U*/;
run;
```

When we apply this technique, we see that after adjustment for the PS, the impact of additional intravascular imaging on TLF occurrence up to 1 year is not statistically significant anymore (p = 0.75).

Number of Observations Read	35389
Number of Observations Used	35389

TLF1Y	Y	1135
	N	34254

Effect	DF	Wald Chi-Square	Pr > Chi-Square	Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio
Intercept	1	9484.3451	<.0001			
IMAGE Y vs N	1	0.0978	0.7545	1.041	0.808	1.342
pred	1	149.9468	<.0001	89.447	43.571	183.626

Figure 12 – Detailed output for Logistic Regression with Adjustment for the Propensity Score

METHOD 2: STRATIFICATION

The second option is to stratify. By using the proc rank statement, we can easily split the studied population in 10 groups based on their PS values.

```
proc rank data=ps groups=10  out=ranked;
var pred;
ranks decile;
run;

proc freq data = ranked ;
tables decile*tlf1y*IMAGE / cmh expected chisq;
run;
```

The treatment effect can be estimated in each stratum, and the estimates can be combined across strata.

When we apply this technique, we see that after stratification for the PS, the impact of additional intravascular imaging on TLF occurrence up to 1 year is not statistically significant anymore (p=0.17).

Summary Statistics for TLF1Y by IMAGE
Controlling for decile

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	1.8599	0.1726
2	Row Mean Scores Differ	1	1.8599	0.1726
3	General Association	1	1.8599	0.1726

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel-Haenszel	1.1813	0.9299	1.5006
	Logit	1.2496	0.9823	1.5897
Relative Risk (Column 1)	Mantel-Haenszel	1.0104	0.9945	1.0266
	Logit	1.0009	0.9894	1.0125
Relative Risk (Column 2)	Mantel-Haenszel	0.8659	0.7055	1.0628
	Logit	0.8234	0.6718	1.0093

Breslow-Day Test for Homogeneity of Odds Ratios	
Chi-Square	8.5373
DF	9
Pr > ChiSq	0.4810

Total Sample Size = 35389

Figure 13 – Detailed output for stratified analysis

METHOD 3: MATCHING 1:X USING GREEDY NEAREST NEIGHBOR MATCHING OPTION

The third option presented here is the matching option. In our specific case, the exposure which is the use of intravascular imaging is only reported by 1642 subjects which represents 5% of our population. So, we are going to look in the remaining 95% of the population for 1642 subject with equal PS value but who did not have the intravascular imaging.

```
proc psmatch data=psm.analysis_image_imputed region=treated;
class   IMAGE  DIAFL BHTENFL BHCHOLFL  PPVDFL PRIMPFL PPTCAFL ACS ITVLM  ITLBIF LGLST LCBOCP ;
psmodel IMAGE(Treated='Y')=
                BBMI DIAFL BHTENFL BHCHOLFL PPVDFL PRIMPFL PPTCAFL ACS ITVLM ITLBIF LGLST LCBOCP ;
        match distance=lps method=greedy(k=1) caliper=1.0
        weight=none;
output out(obs=match)=OutEx matchid=_MatchID;
run;
```

Propensity Score Information

Observations	Treated (IMAGE = Y) N	Mean	Standard Deviation	Minimum	Maximum	Control (IMAGE = N) N	Mean	Standard Deviation	Minimum	Maximum	Treated - Control Mean Difference
All	1642	0.1154	0.1190	0.0022	0.8600	33747	0.0430	0.0468	0.0002	0.7783	0.0724
Region	1642	0.1154	0.1190	0.0022	0.8600	33674	0.0431	0.0468	0.0017	0.7783	0.0723
Matched	1642	0.1154	0.1190	0.0022	0.8600	1642	0.1136	0.1118	0.0022	0.7783	0.0018

Figure 14 – Detailed output for PS Matching analysis

For some subjects, exact matches will be found, however for subjects with extreme values it becomes more challenging to find an exact match. Therefore, you can define the value of a caliper. A caliper is the maximum tolerated difference between matched subjects in a "non-perfect" matching intention. This width is generally defined as a fraction of the standard deviation of the propensity score.

	USUBJID	DIAFL	BHCHOLFL	BHTENFL	BBMI	PPVDFL	PPTCAFL	PRIMPFL	ACS	LCBOCP	LGLST	ITVLM	ITLBIF	IMAGE	_PS_	_MATCHWGT_	_MatchID
1	eU-T123E2_EU-602-011	Y	Y	Y	15.429803688	N	Y	Y	N	Y	N	Y	Y	Y	0.8599607022	1	1
2	eU-T123E2_EU-653-546	Y	Y	Y	19.591836735	Y	Y	Y	N	Y	Y	Y	Y	N	0.7782897588	1	1
3	eU-T123E2_EU-583-005	Y	Y	Y	20.661157025	N	Y	Y	N	Y	Y	Y	Y	Y	0.794763607	1	2
4	eU-T123E2_EU-654-174	Y	Y	Y	25.951557093	N	Y	Y	N	Y	Y	Y	Y	N	0.6578376223	1	2
5	eU-T123E2_EU-603-120	N	N	Y	15.662554078	N	Y	Y	N	Y	Y	Y	Y	Y	0.7660260729	1	3
6	eU-T123E2_EU-653-152	Y	Y	N	24.609375	N	Y	Y	N	Y	N	Y	Y	N	0.6048353359	1	3
7	eU-T123E2_EU-307-123	Y	Y	Y	21.484375	N	Y	Y	Y	Y	N	Y	Y	N	0.6021344685	1	4
8	eU-T123E2_EU-594-013	Y	Y	Y	20.904195011	N	N	Y	N	Y	Y	Y	Y	Y	0.7425853559	1	4
9	eU-T123E2_EU-307-159	Y	Y	Y	24.221453287	Y	Y	Y	N	Y	N	Y	Y	N	0.6014865118	1	5
10	eU-T123E2_EU-595-006	Y	Y	Y	21.203155819	Y	Y	Y	N	Y	Y	Y	Y	Y	0.739324744	1	5
	...																
3275	eU-T123E2_EU-029-353	N	N	N	32.979591837	N	N	N	Y	N	N	N	N	N	0.0041648126	1	1638
3276	eU-T123E2_EU-057-116	N	N	N	34.729638859	N	N	N	Y	N	Y	N	N	Y	0.0041635108	1	1638
3277	eU-T123E2_EU-028-072	N	N	N	33.376654064	N	N	N	Y	N	N	N	N	N	0.0039524335	1	1639
3278	eU-T123E2_EU-436-080	N	N	N	41.617122473	N	N	N	Y	Y	N	N	Y	Y	0.0039517039	1	1639
3279	eU-T123E2_EU-049-040	N	Y	Y	51.132213294	N	Y	N	N	N	Y	N	Y	N	0.0028422489	1	1640
3280	eU-T123E2_EU-066-014	N	N	Y	37.202380952	N	N	N	Y	N	N	N	N	Y	0.0028389824	1	1640
3281	eU-T123E2_EU-066-116	N	N	Y	40.562466198	N	N	N	Y	N	Y	N	N	Y	0.0022946522	1	1641
3282	eU-T123E2_EU-852-193	N	N	N	39.246467818	N	N	N	Y	N	Y	N	N	N	0.0022942597	1	1641
3283	eU-T123E2_EU-028-170	N	N	Y	39.06719835	N	N	N	Y	N	N	N	N	N	0.0022194277	1	1642
3284	eU-T123E2_EU-436-061	N	Y	Y	42.567322787	N	N	N	Y	N	N	N	N	Y	0.002221769	1	1642

Figure 15 – Detailed view of the dataset after PS Matching

In the print screen you can see that each matched pair of exposed and control gets a unique ID, from the highest lower score until the lowest PS score met in the exposure group.

Now that subjects have been matched, let's see if their imbalances are still there or if it has been reduced.

In Figure 16, you can see for each of the 12 predictors the standardized mean differences before and after the matching procedure, marked with a blue cross and a green circle respectively. The blue shade area defines the negligible difference range. In there we can see that after matching for all of the 12 predictors, the imbalance observed between the exposed and control groups have been reduced to negligible.

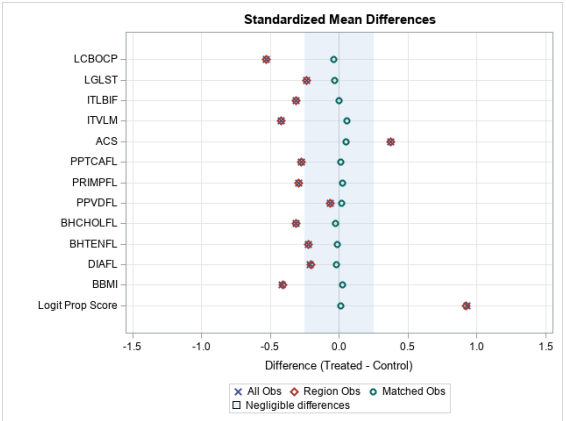


Figure 16 – Standardized Mean Differences after PS Matching

	Unmatched				Matched		
	Treated (IMAGE='Y')	Control (IMAGE='N')	p-value		Treated (IMAGE='Y')	Control (IMAGE='N')	p-value
	N = 1642	N = 33 747			N = 1642	N = 1642	
AGE	66.2 ± 11.23	64.18 ± 11.22	<.0001		66.2 ± 11.23	66.9 ± 11.37	0.10
BBMI	26.1 ± 4.51	27.9 ± 4.13	<.0001		26.1 ± 4.51	26.00 ± 3.88	0.61
DIAFL	37.09%	27.35%	<.0001		37.09%	36.36%	0.66
BHTENFL	71.25%	60.62%	<.0001		71.25%	70.40%	0.59
BHCHOLFL	66.50%	51.25%	<.0001		66.50%	65.10%	0.40
PPVDFL	7.73%	5.99%	0.0037		7.73%	8.22%	0.61
PRIMPFL	15.71%	6.45%	<.0001		15.71%	16.69%	0.45
PPTCAFL	36.24%	23.74%	<.0001		36.24%	36.54%	0.86
ACS	37.39%	55.77%	<.0001		37.39%	39.89%	0.14
ITVLM	14.01%	2.63%	<.0001		14.01%	15.59%	0.20
ITLBIF	23.20%	11.41%	<.0001		23.20%	23.08%	0.93
LGLST	48.78%	37.19%	<.0001		48.78%	47.14%	0.35
LCBOCP	70.16%	44.83%	<.0001		70.16%	68.27%	0.24

Figure 17 – Comparison of Distribution before and after PS Matching

If you want to have a closer look for each variable, you can get a boxplot presenting the numerical distribution per treatment group before and after matching, such as for example for the BMI, or a stacked bar presenting proportional distribution for dichotomous categorical variable such as a flag.

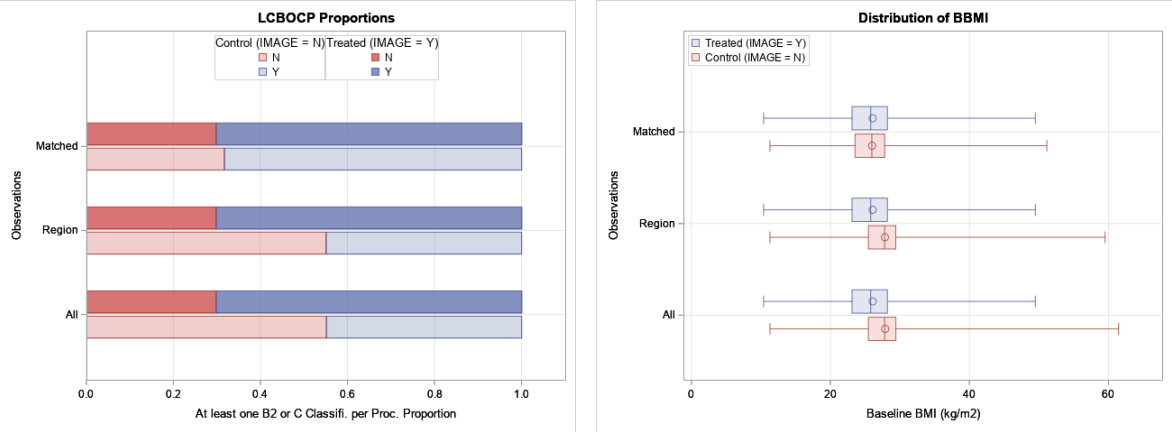


Figure 18 – Standardized Mean Differences after PS Matching

Now that we have the paired subjects and verified that their covariates are distributed similarly, let's see how the treatment impacts the outcome. To do so, you need to merge your dataset output containing the PS values and Matched ID with the outcome variable. As a reminder, we are only using 10% of the population, as exposure was observed in 5% and for each exposed subject, we selected one control subject, off course without taking the outcome into account.

```
proc sort data=OutEx ; by usubjid ; run ;
data  OutExTLF  ;
merge   OutEx
        psm.analysis_image_imputed (keep = usubjid tlf1y);
by usubjid ;
run ;

proc freq data=OutExTLF ;
table tlf1y*image / nopercent norow chisq ;
run;
```

The FREQ Procedure

Frequency Col Pct	Table of TLF1Y by IMAGE		
TLF1Y(Target Lesion Failure up to 1Y)	IMAGE(Imaging Performed at Index Procedure)		
	N	Y	Total
N	1557 94.82	1559 94.95	3116
Y	85 5.18	83 5.05	168
Total	1642	1642	3284
Frequency Missing = 32105			

When we apply this technique, we see that after matching the subjects based on their PS values, the impact of additional intravascular imaging on TLF occurrence up to 1 year is not statistically significant anymore (p=0.87).

This method was very popular in the last 10 years but has some limitations. For example, here, the analysis only considers 10% of the subjects. We tried to increase it by matching not with a 1:1 ratio but with a 1:5 or 1:10 ratio and see if the differences was still negligible and if the impact of the exposure on the outcome would remain non statistically significant – and it did.

Statistics for Table of TLF1Y by IMAGE

Statistic	DF	Value	Prob
Chi-Square	1	0.0251	0.8741
Likelihood Ratio Chi-Square	1	0.0251	0.8741
Continuity Adj. Chi-Square	1	0.0063	0.9369
Mantel-Haenszel Chi-Square	1	0.0251	0.8742
Phi Coefficient		-0.0028	
Contingency Coefficient		0.0028	
Cramer's V		-0.0028	

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	0.9752	0.7149	1.3303
Relative Risk (Column 1)	0.9876	0.8471	1.1515
Relative Risk (Column 2)	1.0127	0.8656	1.1848

Sample Size = 3284
Frequency Missing = 32105

WARNING: 91% of the data are missing.

Figure 19 – Detailed output after PS Matching and impact on outcome analysis

METHOD 4: ADJUSTMENT BY INVERSE PROBABILITY OF TREATMENT WEIGHTING (IPTW)

The last method presented is the adjustment by inverse probability of treatment weighting. This technique is now used the most in publication. Starting from the PS generated, you calculate a weight depending on the treatment group attribution.

```
%let class = DIAFL BHTENFL BHCHOLFL PPVDFL PRIMPFL PPTCAFL ACS ITVLM ITLBIF LGLST LCBOCP ;
%let cont = BBMI ;

data ips;
   set  ps ;
   if image = "N" then weight1 = 1/(1-pred);
   if image = "Y" then weight1 = 1/(pred);
run;

proc univariate data=ips ;
var weight1 ;
run ;
```

To avoid the effect of outlier values on your analysis, you can decide to remove subjects from the upper and lower 1% of your distribution (% to be excluded depends on your initial sample size and the PS distribution in both groups).

```
data ips_use;
   set ips;
   if weight1 > 32.49160 then delete; *Exclude 1% upper cases;
   if weight1 < 1.00375 then delete;  *Exclude 1% lower cases;
   logit_ps = log(pred/(1-pred));
run;
```

Then you will use this weight in your analysis to estimate the effect of the treatment on your outcome.

```
proc causaltrt data = ips_use METHOD=IPWR ATT;
   class image tlf1y  &class;
   psmodel image (REF="N") = &class &cont;
   model tlf1y = /  DIST=BIN;
   ods output   CausalEffects= CausalEffects_ATT_IPWR;
run;
```

IPTW model (ATT)

The CAUSALTRT Procedure

Model Information	
Data Set	WORK.IPS_USE
Distribution	Binomial
Estimation Method	IPWR
Treatment Variable	IMAGE
Outcome Variable	TLF1Y

Number of Observations Read	34681
Number of Observations Used	34681

```
data CausalEffects_ATT_IPWR;
   set CausalEffects_ATT_IPWR;
   /*CALCULATE ESTIMATED ODDS RATIO
        AND CONFIDENCE INTERVAL*/
   OR= exp(estimate);
   OR_LOW= exp(lowerwaldcl);
   OR_UP= exp(upperwaldcl);
run;
```

Here, once again, when we apply this technique, we see that after adjustment by weighting of the subjects based on their PS values, the impact of additional intravascular imaging on TLF occurrence up to 1 year is not statistically significant anymore (p=0.82).

Response Profile

Ordered Value	TLF1Y	Total Frequency
1	Y	1114
2	N	33567

PROC CAUSALTRT is modeling the probability that TLF1Y='Y'.

Treatment Profile

Ordered Value	IMAGE	Total Frequency
1	Y	1288
2	N	33393

	Parameter	Level	OR	OR_LOW	OR_UP	ProbZ
1	POM	Y	1.0550323485	1.0421380139	1.0680862243	<.0001
2	POM	N	1.056997387	1.0462088516	1.0678971742	<.0001
3	ATT		0.998140924	0.9825650733	1.0139636867	0.8166

Figure 20 – Detailed output after IPTW and impact on outcome analysis

CONCLUSION

Now that we have seen in detail these 4 options for analysis, let's take a step back in order to get a better view of the full picture in a forest plot. In the crude analysis, the use of intravascular imaging was increasing the risk of TLF up to 1 year, and the OR was equal to 1.6 with 95% confidence intervals ranging from 1.3 to 2.1. This difference was statistically significant.

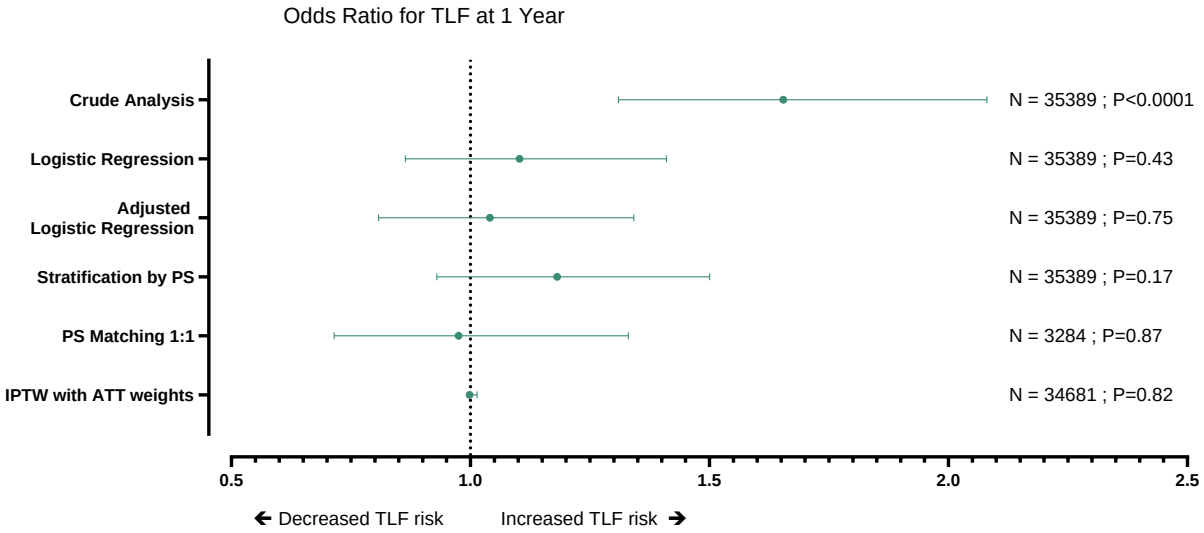


Figure 21 – Forest Plot of the crude analysis and the 4 methods of PSM

Once we adjusted the impact of the exposure on the outcome for the identified covariates, all analysis options gave another result, where the impact of additional intravascular imaging on TLF occurrence up to 1 year is not statistically significant anymore.

Each technique has its advantages and limitations, but they are all confirming the same findings, and the 95% Confidence intervals are quite broad, except for the last one, where the interval is very small.

As a conclusion, propensity score analysis have some great advantages and allow the observational studies to gain in clinical evidence when compared to randomized clinical trials. The first one is that you can include many confounders when you calculate your PS values and by doing so you separate the confounder adjustment from the outcome analysis. Finally, this analysis procedure has a multistep process where multiple alternatives and options can be used to fit your needs.

Off course such as all analysis techniques, there are some limitations, and these should be reminded when interpreting the results. The first one is that we can only use what has been collected. Even if you adjust for every variable, there will always be a possible factor that impacts the outcome but has not been collected as it was not yet considered as a possible predictor at the time of collection. The second one is that this technique requires some statistical expertise to properly define and validate every step while PS analysis is not a well-known technique (yet). Finally, when applying PS analysis, there is no estimation of the effects of the confounders on the outcome, neither on their interactions.

REFERENCES

eUltimaster Terumo Page : https://www.terumo-europe.com/en-emea/clinicaldata/e-ultimaster-trial

Rosenbaum PR, Rubin DB. *The central role of the propensity score in observational studies for causal effects*. Biometrika. 1983; 70:41–55.

SAS Institute Inc. *Base SAS® 9.4 Procedures Guide, Seventh Edition; 2017.* Cary, NC, USA.

SAS Institute Inc. *SAS/STAT 15.1 User's Guide, The PSMATCH Procedure; 2018.* Cary, NC, USA.

Austin PC. *The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments*. Stat Med. 2014 Mar 30; 33(7):1242-58.

ACKNOWLEDGMENTS

I wish to thank all former and current colleagues who generously provided their support, insight and perspective that greatly helped to develop the presentation and improve the quality of this paper, but in particular: Lotte Jacobs, Frank Van Leeuwen, Laurence Seidel, Erik Spaepen, Natalie Bolton and Vladimir Borovicanin.

CONTACT INFORMATION

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

DEBRUS Roxane
TERUMO EUROPE NV.
Researchpark Haasrode 1520
Interleuvenlaan 40
3001 Leuven, Belgium
Work Phone: +32 16 38 12 81
Email: roxane.debrus@terumo-europe.com
Web: https://www.terumo-europe.com/en-EMEA