Implementation of Propensity Score Analysis in Real World Evidence Studies Based on CDISC Standards

Presented by Sisi Zhou, Senior Statistical Programmer,
Statistical Programming, Parexel

Meet the Speaker

Sisi Zhou

Title: Senior Statistical Programmer

Organization: Parexel

Sisi Zhou is a senior statistical programmer, with over six years' experience in pharmaceutical industry. She has rich experience in applying emerging statistical and programming methods in the data analysis and reporting. She is well skilled in SAS and R, and developed several SAS macros and R functions to generate data packages following CDISC standards.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

Agenda

1. Background
2. Propensity Score Analysis Dataset (ADPS)
3. Steps of propensity score analysis
4. Reference

Background

Real World Evidence (RWE) studies

- RWE studies collect real-world data from routine healthcare practice.

- RWE studies provide a large picture on how a treatment is used and how it performs under real-world conditions.

- Most RWE studies are observational in nature.

Limitation of observational RWE studies

- Baseline confounding factors potentially may be identified in RWE studies due to the observational nature.
- Let's see an example of Simpson's Paradox:

Table: Success rate in removing kidney stones by treatment method*

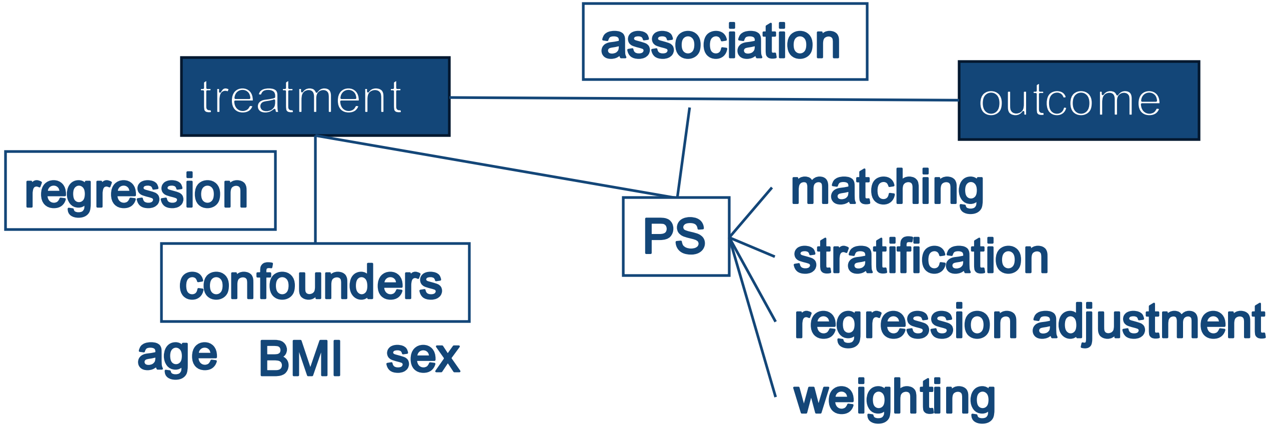
Treatment	Severity of illness		
	Small stones (Success Rate %)	Large stones (Success Rate %)	Total (Success Rate %)
A	81/87 (93)	192/263 (73)	273/350 (78)
B	234/270 (87)	55/80 (69)	289/350 (83)

* Data from Charig [1]

Propensity score (PS) analysis

- PS analysis attempts to estimate the effect of a treatment by accounting for the covariates that predict receiving the treatment.
- The goal is to balance the distribution of covariates between treatment.

Figure: Propensity score in the relationship of treatment and outcome



Techniques to use propensity score

- Matching
  - A subject is randomly selected from one treatment group and matched with a subject from the other treatment group based on the propensity score.
- Stratification
  - Group subjects into subsets based on the propensity score and compare the subjects from different treatment groups within the same subset.
- Regression adjustment
  - Add the propensity score in the regression to adjust the covariate imbalances.
- Weighting
  - Weight the subjects from different treatment groups by propensity score.

Confounder selection

- Select a group of potential confounders based on the historical research on the treatment and outcomes.
- Identify the actual confounders with different distributions between treatment groups via descriptive statistics and regression models.

Propensity Score Analysis Dataset (ADPS)

ADPS elements

- Treatment
  - Come from ADSL.
- Outcomes
  - Various, may come from several ADaM datasets.
- Potential confounders
  - Baseline characteristics from ADSL.
- Confounder selection flags – to identify the actual confounders from all potential confounders
  - Select the confounders with different distributions between treatment groups.
  - May consider the number of missing values of the confounders when selecting the confounders.

ADPS structure

Basic Data Structure (BDS)

- One record per subject, per analysis parameter, per analysis timepoint
- Outcome as row (PARAM)
- Confounder as column (Variable)
- Confounder selection flag as column (Variable)

PARAM	Confounder A	Conf A Select	Confounder B	Conf B Select
Outcome 1	33	Y	Never	N
Outcome 2	44	N	Current	Y
Outcome 3	55	Y	Past	N

ADPS variables

Variables come from ADSL

- Identifier Variables
  - STUDYID, USUBJID, SUBJID, SITEID…
- Subject Demographics Variables
  - AGE, SEX, RACE…
- Population Indicator Variables
  - FASFL, SAFFL, ITTFL, PPROTFL…
- Treatment Variables
  - ARM, ACTARM, TRTxxP, TRTxxPN, TRTxxA, TRTxxAN…
- Treatment Timing Variables
  - TRTSDT, TRTSTM, TRTSDTM, TRTEDT, TRTETM, TRTEDTM, TRxxSDT, TRxxEDT…

ADPS variables

Variables defined in ADPS:

- **Analysis Parameter Variables**

Variable Name	Variable Label	Type	Comment
PARAMCD	Parameter Code	Char	Outcomes
PARAM	Parameter	Char	
PARAMN	Parameter (N)	Num	
AVAL	Analysis Value	Num	
AVALC	Analysis Value (C)	Char	
AVALCATy	Analysis Value Category y	Char	
AVALCAyN	Analysis Value Category y (N)	Num	

ADPS variables

Variables defined in ADPS:

- **Analysis Parameter Variables**

Variable Name	Variable Label	Type	Comment
BASE	Baseline Value	Num	
BASEC	Baseline Value (C)	Char	
BASECATy	Baseline Category y	Char	
BASECAyN	Baseline Category y (N)	Num	

ADPS variables

Variables defined in ADPS:

- **Analysis Descriptor Variables**

Variable Name	Variable Label	Type	Codelist	Core	Comment
DTYPE	Derivation Type	Char	(DTYPE)	Cond	Analysis value derivation method.

ADPS variables

Variables defined in ADPS:

- **Time-to-Event Variables (related to Time-to-Event outcomes)**

Variable Name	Variable Label	Type	Codelist	Core	Comment
STARTDT	Time-to-Event Origin Date for Subject	Num		Perm	
STARTDTM	Time-to-Event Origin Datetime	Num		Perm	
STARTDTF	Origin Date Imputation Flag	Char	(DATEFL)	Cond	
STARTTMF	Origin Time Imputation Flag	Num		Cond	
CNSR	Censor	Num		Cond	
EVNTDESC	Event or Censoring Description	Char		Perm	
CNSDTDSC	Censor Date Description	Char		Perm	

ADPS variables

Variables defined in ADPS:

- **Population Indicator Variables**

Variable Name	Variable Label	Type	Codelist	Core	Comment
SUBGRP	Subgroup	Char		Cond	Per analysis needs.

ADPS variables

Variables defined in ADPS:

- **Analysis Flag Variables**

Variable Name	Variable Label	Type	Codelist	Core	Comment
ANL01FL	Analysis Flag 01	Char	Y	Cond	Flag the outcome values that meet the PSA inclusion criteria.
ANL02FL	Analysis Flag 02	Char	Y	Cond	Flag the records used in calculating propensity score.
ANLzzFL	Analysis Flag zz	Char	Y	Cond	Per analysis needs.

ADPS variables

Variables defined in ADPS:

- **Confounder Variables**

Variable Name	Variable Label	Type	Core	Comment
<CONFOUNDER> e.g. SMOKE	<Confounder Label>	Char/Num	Perm	Potential confounders

ADPS variables

Variables defined in ADPS:

- **Confounder Selection Flag Variables**

Variable Name	Variable Label	Type	Codelist	Core	Comment
<CONFOUNDER>YN e.g. SMOKEYN	Confounder Selected Flag	Char	Y, N	Perm	

Steps of propensity score analysis

Steps of propensity score analysis

1. Define the treatment, outcome, and potential confounders.
2. Assess the balance of the potential confounders between treatment groups and identify the actual confounders.
3. Estimate the propensity score based on the actual confounders.
4. Re-assess the balance of the confounders between treatment groups by adding the propensity score as one covariate.
5. Analyze the treatment effect on the outcome by adjusting for the propensity score.
6. Perform sensitivity analysis as needed.

Output: balance of the potential confounders between treatment groups

	Balance Assessment[a]	
	Odds Ratio	P-value
Potential Confounders		
Age	1.02	0.413
BMI	0.78	0.003
Sex		
Male	42.25	<0.001
Female		

[a] A logistic regression model with treatment as dependent variable. The independent effect include each potential confounder separately.

Note: The data used in the analysis were dummy data.

Program: balance of the potential confounders between treatment groups

```
proc logistic data = adps;
    class treatment sex;
    model treatment = sex;
    oddsratio sex / diff = ref;
    ods output OddsRatios = OR
                        ParameterEstimates = ESTIMATE;
run;
```

Output: propensity score category

	Treatment A	Treatment B
Estimated Propensity Score Class		
N		
0% ~ 20%	0	5
20% ~ 40%	1	6
40% ~ 60%	3	3
60% ~ 80%	5	0
80% ~ 100%	6	1

Note: The data used in the analysis were dummy data.

Program: propensity score calculation

```
proc logistic data = adps;
    class treatment sex;
    model treatment = BMI sex;
    output out = PScore predicted = ps;
run;
```

Output: unadjusted and adjusted balance of the confounders between treatment groups

	Unadjusted Analysis[a]		Adjusted Analysis[b]	
	Odds Ratio	P-value	Odds Ratio	P-value
Potential Confounders				
Age	1.02	0.413		
BMI	0.78	0.003	1.01	0.932
Sex				
Male	42.25	<0.001	4.00	0.424
Female				

[a] A logistic regression model with treatment as dependent variable. The independent effect include each potential confounder separately.

[b] A logistic regression model with treatment as dependent variable. The independent effects include each potential confounder and the estimated PS in five classes.

Note: The data used in the analysis were dummy data.

Program: adjusted balance of the confounders between treatment groups

```
proc logistic data = adps;
    class treatment sex pscl;
    model treatment = sex pscl;
    oddsratio sex / diff = ref;
    ods output OddsRatios = OR_adj
                    ParameterEstimates = ESTIMATE_adj;
run;
```

Output: unadjusted and adjusted treatment effect on the outcome

	Unadjusted Analysis[a]		Adjusted Analysis[b]	
	Odds Ratio	P-value	Odds Ratio	P-value
Treatment A vs B	0.13	0.014	<0.01	0.892

[a] A logistic regression model with binary outcome as dependent variable. The independent effect include treatment.
[b] A logistic regression model with binary outcome as dependent variable. The independent effects include treatment and the estimated PS in five classes.
Note: The data used in the analysis were dummy data.

Program: adjusted treatment effect on the outcome

```
proc logistic data = adps;
    class outcome treatment pscl;
    model outcome = treatment pscl;
    oddsratio treatment / diff = ref;
    ods output OddsRatios = OR_adj_out
                        ParameterEstimates = ESTIMATE_adj_out;
run;
```

Reference

Reference

[1] Charig C.R., Webb D.R., Payne S.R., Wickham J.E. Comparison of Treatment of Renal Calculi by Open Surgery, Percutaneous Nephrolithotomy, and Extracorporeal Shockwave Lithotripsy. Br. Med. J. 1986;292:879–882.

Thank You!

Contact:

Email: Sisi.Zhou@parexel.com