Propensity Score Analysis

- Bhagyashree Patil & Garima Joshi

03 December, 2016

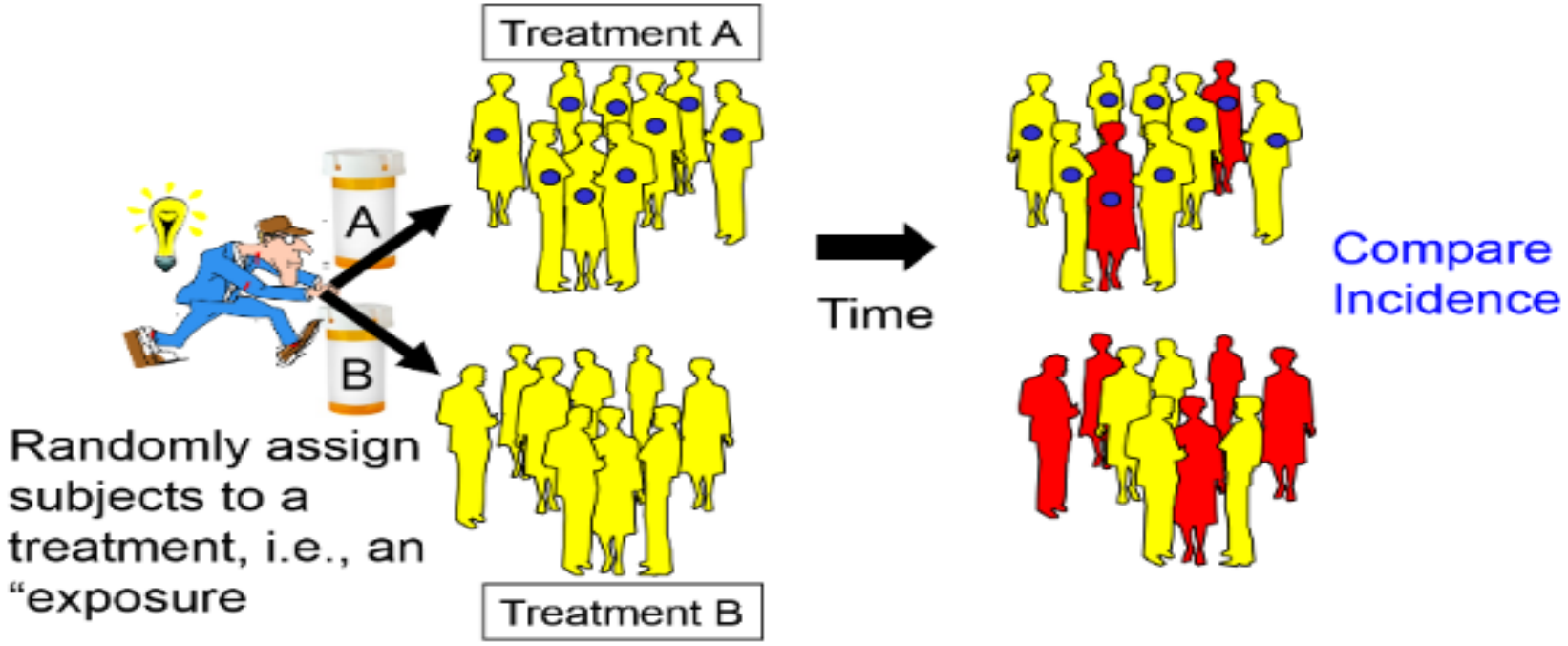
sciformix
Trusted Services. Built on Science.

Table of Contents

sciformix
Trusted Services. Built on Science.

- Randomized Trials and Observational Studies
- Propensity Scores
- Conditioning Methods
- Nearest Neighbor matching
- Examples
- Conclusion

Randomized Trials



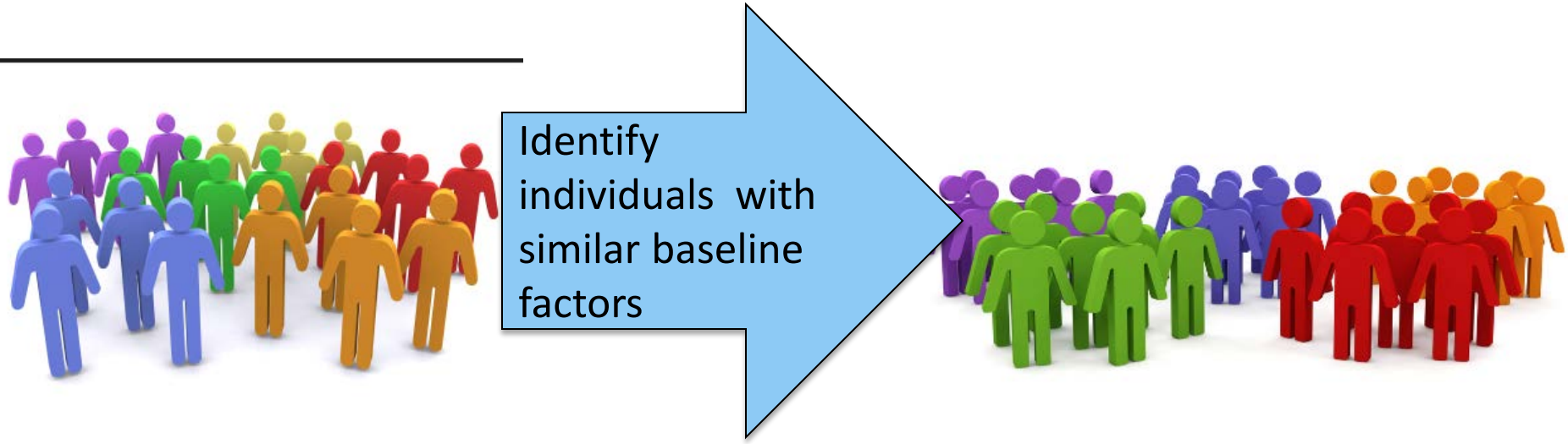
Enrollment
- Inclusion
- Exclusion

Treatment Allocation
- Randomization
- Ensure compliance

Analysis

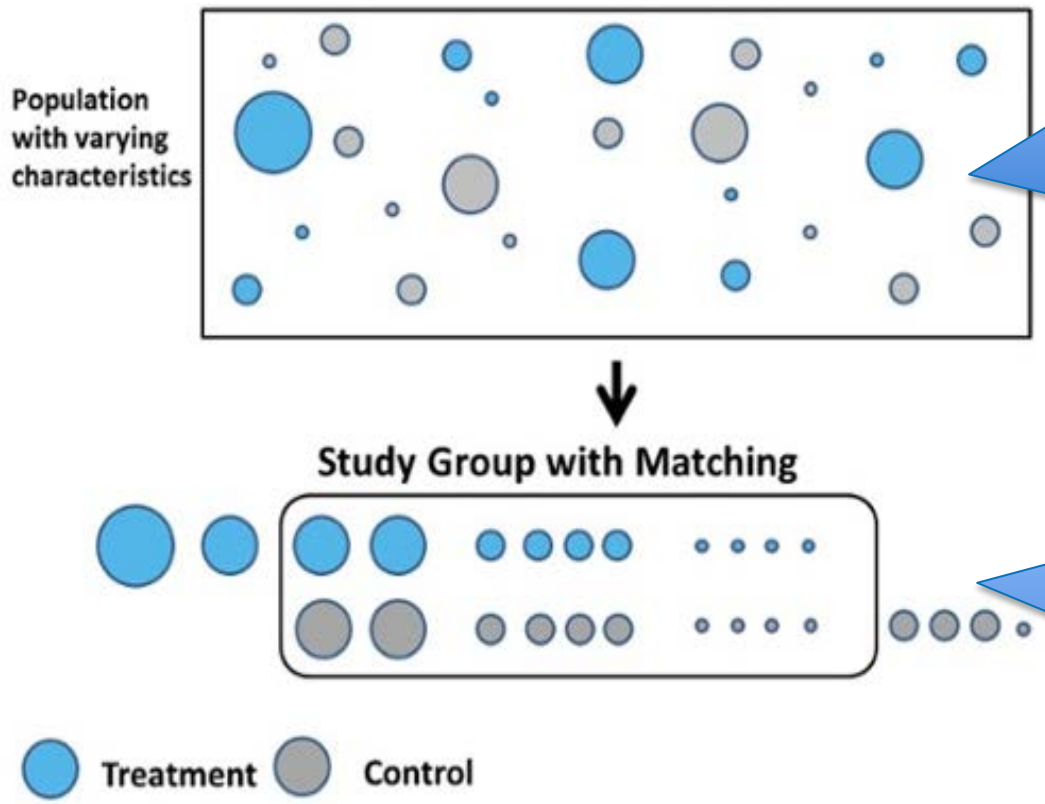
Observational studies

- Heterogeneity among patients- prognostic factors, demographics, etc
- Compliance to doses/procedures cannot be ensured
- Presence of selection bias



Propensity Score

- The propensity score was defined by Rosenbaum and Rubin (1983) to be the probability of treatment assignment conditional on observed baseline covariates.



The probability of being in the treatment group is estimated for each subject in the sample, given their observable characteristics

The treatment effect is estimated as the difference between the treatment groups and their statistically-matched control groups.

Propensity Score Cont....

- Propensity score model assumes that all variables that affect treatment and outcome have been measured.
- The propensity score is most often estimated using a logistic regression model, in which treatment status is regressed on observed baseline characteristics.

The propensity scores for the treatment variable GROUP, predicted from three covariates (var1 – var3).

```
proc logistic  data = ;
     model group = var1-var3/lackfit outroc = ps_r;
     output out= ps_p XBETA=ps_xb STDXBETA= ps_sdxb
    PREDICTED = ps_pred;
run;
```

Choosing variables for PS

- Include : Theoretically related to treatment & outcome
- Include : Available & easy/reliable to collect on everyone
- Include : Correlated with unmeasured confounders

- Do not include : Variables that may be affected by the treatment
- Do not include : Variables that predict treatment status perfectly

Common variables in PS

- Demographics (age, gender, ethnicity/race, marital status)
- Illness-related factors (primary diagnosis, comorbid conditions, severity of illness)
- Setting (urban/rural, geographic region, hospital site/type)
- Clinician characteristics (yrs in practice, specialty)

What PS can & cannot do

- Help find matches from comparison group so that measured confounders can be equally distributed between treatment & comparison groups
- Helps improve precision of estimates of treatment effects
- Cannot account for unmeasured confounders – only control for observed variables and only to the extent that they are accurately measured

Conditioning Methods

- Conditioning methods attempt to replicate features of randomized experiments, like create groups that look only randomly different from one another (at least on observed variables) .

- Different ways to use propensity scores to correct estimates of treatment effect for selection bias include-

➢ Nearest Neighbor Matching on propensity score

➢ Stratification based on propensity score

➢ Using propensity score as a covariate

- Once optimal balance is obtained between the two treatment groups on the observed covariates, treatment effects can be analyzed.

Nearest Neighbor Matching

- NN Matching selects a control unit for each treated unit based on the smallest distance from that treated unit in Propensity score.

For example, for the i^{th} patient in trt A and the j^{th} patient in trt B, the two are said to be matched if the difference in the propensity scores is-

$$abs(PS_i - PS_j) <= \delta$$

δ- caliper width(pre-determined value)

- This selection process can be done with/without replacement, i.e., subjects are may/may not return to the sample after being pair-matched.

Algorithm for matching

- Choose a subject from treatment A
- Find the closest subject in treatment B.
- If the distance between subject from treatment A and from treatment B is acceptable, record the match.
- Without replacement-
  - Remove this subject from treatment A from the list of subjects from treatment A.
  - Remove this subject from treatment B from the list of subjects from treatment B.
- Go back to step 1.

Challenges of Nearest Neighbor Matching

Since subjects are not returned to the sample after being pair-matched-

- Many of the subjects in the data are discarded.
- The final estimates depend on the order in which the observations are matched. To avoid this we can randomly order the sample before matching.
- If the caliper size is too large there is a risk that very dissimilar individuals will be matched, while if the caliper size is too small the sample size may become too small to obtain statistically convincing results.

Example

- An Observational study consisting of 440 Patients suffering Rheumatoid arthritis.
- Treatment groups : A and B
- We want to see the effect of treatments on the ACR50 score(binary variable) after 6 months.
- ACR50 is a scale to measure change in rheumatoid arthritis symptoms. If ACR50=1 it means RA has improved by 50%.

Observed baseline characteristics

Demographics
Age
Gender

Illness Related Factors
Previously Failed Biologic DMARDS(Disease-modifying antirheumatic drugs)
Duration of RA

- Fit a logistic regression model where treatment is regressed on these baseline characteristics to get the propensity score.

- Nearest Neighbor Matching on propensity score

Data Summary

The percentage of subjects in each treatment group for categorical covariates

	Before Matching	
	A	B
Failed Prior DMARD	12.86	2.01
Gender(Male)	14.94	21.11

Mean of subjects in each treatment group for continuous covariates

	Before Matching	
	A	B
Age(Mean)	57.7	49.9
Disease duration(Mean)	8.1	8.2

Results (Before matching the data)

Multivariate logistic regression analysis to evaluate the odds of not achieving ACR50 -

Odds ratio	Estimate	Lower CI	Upper CI	P-Value
Age (units=12)	1.240	1.017	1.511	0.0338
SEX (Male vs Female)	0.955	0.571	1.596	0.8596
Disease duration(units=8)	1.584	1.264	1.984	<.0001
treatment A vs B	1.590	1.051	2.404	0.0280

-odds of not achieving ACR50 has increased by 24% with 12 unit increase in age .

-odds of not achieving ACR50 is 1.5 times higher in treatment group A than in treatment group B.

Results (After matching the data)

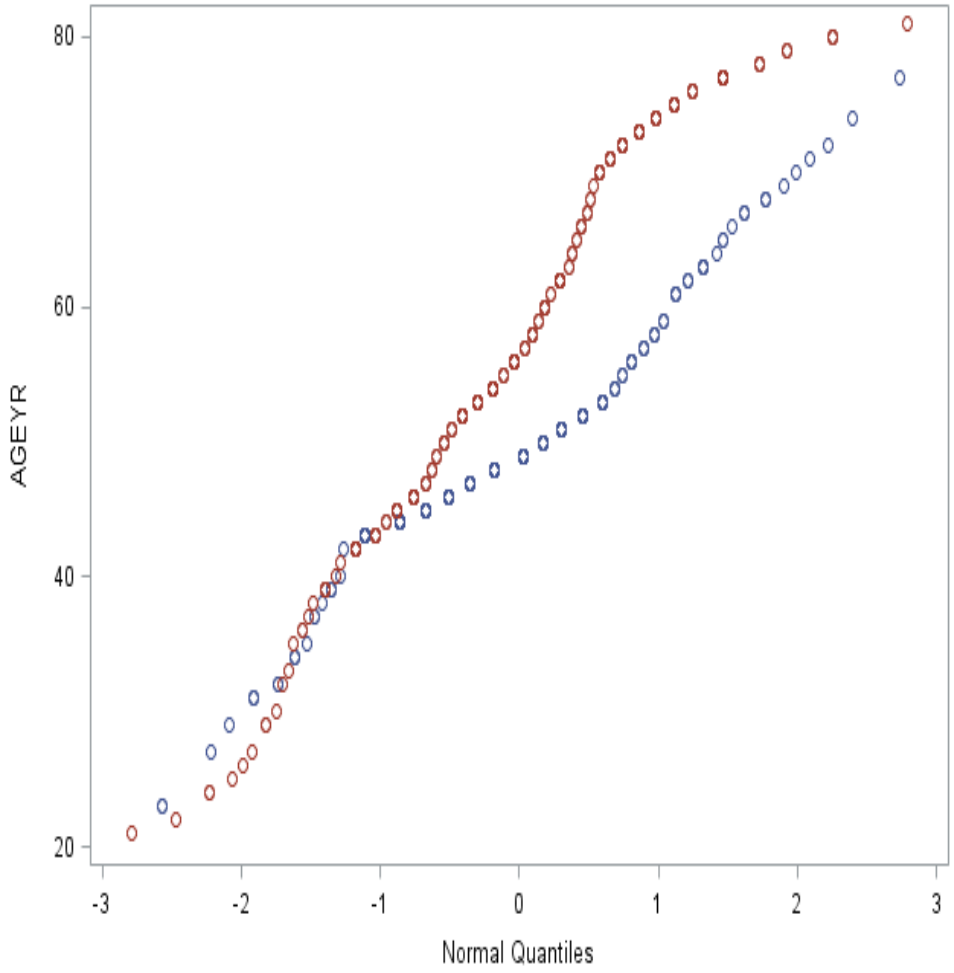
1)Out of 440 observations , 380 observations were matched (δ=0.1).

Odds ratio	Estimate	Lower CI	Upper CI	P-Value
Age (units=12)	1.141	0.911	1.430	0.2513
SEX (Male vs Female)	1.038	0.606	1.777	0.8920
Disease duration(units=8)	1.602	1.260	2.036	0.0001
treatment A vs B	1.514	0.984	2.329	0.0593

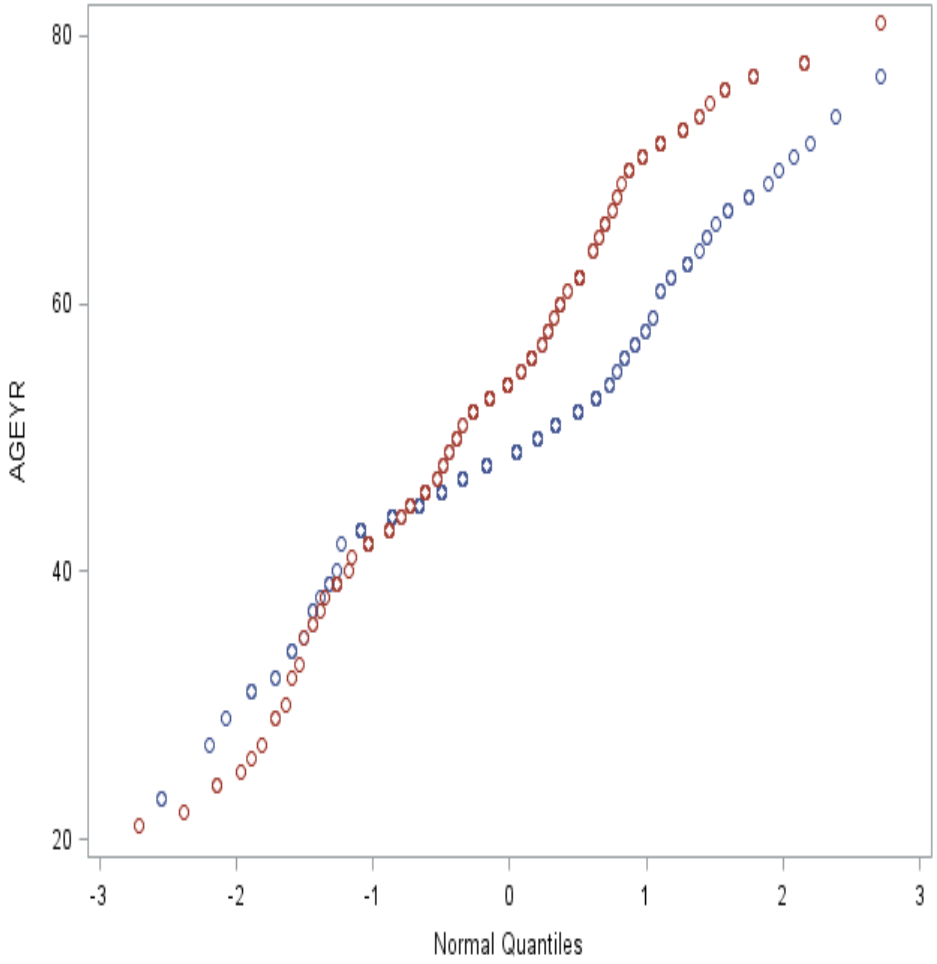
-As age is balanced between the treatment groups, its effect on the treatment is removed and treatment difference is no longer significant.

QQ plot for Age

Before matching



After matching



Treatment A Treatment B

Conclusion

- Appropriate implementation of propensity score adjustments is a multi-step process presenting many alternatives in terms of estimation and conditioning methods.

- Useful when adjusting for a large number of risk factors

- Deep understanding of the study is required by the researcher for selecting observable covariates to include in calculating the propensity score.

- Evidence of imbalance between covariates across the two groups may require re-estimation of PS using a more elaborate model (with interaction effects or polynomial terms).

References

- Austin PC.; An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res.;2011

- Propensity score analysis and assessment of propensity score approaches using sas® procedures; Rheta E. Lanehart, Patricia Rodriguez de Gil, Eun Sook Kim, Aarti P. Bellara, Jeffrey D. Kromrey, and Reginald S. Lee, University of South Florida

sciformix
Trusted Services. Built on Science.

Thank you