

A Practical Introduction to Propensity Score Analysis using R

Ehsan Karim [<http://ehsank.com/>]

30 Sept 2020: SPPH, UBC

About this event

- ▶ TI Methods Speaker Series page: Therapeutics Initiative
 - ▶ Dr. Carole Lunny
- ▶ SARGC page: Students and Recent Graduates Committee (SARGC) coordinate activities for the Statistical Society of Canada (SSC)'s student and recent graduate members
 - ▶ Md. Erfanul Hoque
 - ▶ Janie Coulombe

Outline

- ▶ [1] Data and Regression
 - ▶ (Diagnostics)
- ▶ [2] Exact matching
 - ▶ (motivation)
- ▶ [3] Propensity score matching
 - ▶ (4 steps)
- ▶ [4] Propensity score Reviews in different disease areas
 - ▶ (brief)

[1] Right Heart Catheterization (RHC) Dataset

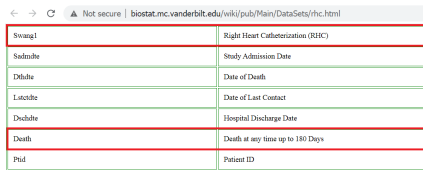
The dataset that we will use today is from Connors et al. (1996).

The effectiveness of right heart catheterization in the initial care of critically ill patients

AF Connors, T Speroff, NV Dawson, C Thomas... - *Jama*, 1996 - [jamanetwork.com](#)

Objective.—To examine the association between the use of right heart catheterization (RHC) during the first 24 hours of care in the intensive care unit (ICU) and subsequent survival, length of stay, intensity of care, and cost of care. Design.—Prospective cohort study. Setting.—Five US teaching hospitals between 1989 and 1994. Subjects.—A total of 5735 critically ill adult patients receiving care in an ICU for 1 of 9 prespecified disease categories. Main Outcome Measures.—Survival time, cost of care, intensity of care, and length of stay in the ...

☆ 99 Cited by 2588 Related articles All 11 versions Import into BibTeX



The screenshot shows a web browser window with the URL bioestat.mc.vanderbilt.edu/wiki/Main/DataSets/rhc.html. The page contains a table with two columns. The first column lists variable names, and the second column lists their descriptions. The variables are: swang1 (Right Heart Catheterization (RHC)), sadmte (Study Admission Date), dthlde (Date of Death), lctcdte (Date of Last Contact), dchdte (Hospital Discharge Date), Death (Death at any time up to 180 Days), and ptiid (Patient ID). The rows for swang1, Death, and ptiid are highlighted with red borders.

Variable	Description
swang1	Right Heart Catheterization (RHC)
sadmte	Study Admission Date
dthlde	Date of Death
lctcdte	Date of Last Contact
dchdte	Hospital Discharge Date
Death	Death at any time up to 180 Days
ptiid	Patient ID

[Frank E Harrell Jr.](#)

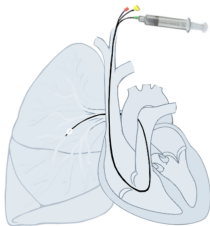
Last modified: Fri Dec 27 16:14:18 EST 2002

Notations

- ▶ Outcome Death (Y)
 - ▶ Death at any time up to 180 Days
- ▶ Treatment swang1 (A : Swan-Ganz catheter)
 - ▶ Whether or not a patient received a RHC
- ▶ Covariate list: L (age, sex, race, ...)
- ▶ Analysis strategy: matching RHC patients with non-RHC patients

[1] Right Heart Catheterization (RHC) Dataset

- ▶ RHC is helpful in guiding therapy decision
 - ▶ Helps determine the pressures within the heart
- ▶ Popularly believed that RHC is beneficial
- ▶ Conducting RCT is hard (ethical reasons)
- ▶ Benefit of RHC was not shown earlier (1996)
- ▶ SUPPORT data has 2 phases
 - ▶ phase 1: prospective observational study
 - ▶ phase 2: cluster RCT
 - ▶ Data in this study is combined



[1] Load data

```
# Load the cleaned up data.  
# Reproducible codes:  
# https://ehsanx.github.io/SARGC-TIMethods/  
analytic.data <- readRDS("data/RHC.Rds")  
# Data size and number of variables  
dim(analytic.data)
```

```
## [1] 4767 23
```

```
# variable names  
names(analytic.data)
```

```
## [1] "age" "sex" "race"  
## [4] "Disease.category" "DNR.status" "APACHE.III.score"  
## [7] "Pr.2mo.survival" "No.of.comorbidity" "DASI.2wk.prior"  
## [10] "Temperature" "Heart.rate" "Blood.pressure"  
## [13] "Respiratory.rate" "WBC.count" "PaO2.by.FIO2"
```

[1] Inspecting data: Crude

```
require(tableone)
# 2 x 2 table
tab0 <- CreateTableOne(vars = "RHC",
                        data = analytic.data,
                        strata = "Death")
print(tab0, showAllLevels = TRUE)
```

```
##           Stratified by Death
##           level  No          Yes          p          test
##    n           2013          2754
##  RHC (%) No RHC 1315 (65.3) 1268 (46.0) <0.001
##           RHC      698 (34.7) 1486 (54.0)
```


[1] Inspecting data: Some baseline variables

```
baselinevars <- c("age", "sex", "race")  
# Table 1  
tab1 <- CreateTableOne(vars = baselinevars,  
                        data = analytic.data,  
                        strata = "Death", includeNA = TRUE,  
                        test = TRUE, smd = FALSE)  
print(tab1, showAllLevels = FALSE, smd = FALSE)
```

##		Stratified by Death			
##		No	Yes	p	test
##	n	2013	2754		
##	age (%)			<0.001	
##	[-Inf,50)	713 (35.4)	400 (14.5)		
##	[50,60)	351 (17.4)	452 (16.4)		
##	[60,70)	426 (21.2)	789 (28.6)		
##	[70,80)	382 (19.0)	750 (27.2)		
##	[80, Inf)	141 (7.0)	363 (13.2)		

[1] Crude regression

```
# adjust the exposure variable (primary interest)
fit0 <- glm(I(Death=="Yes")~RHC,
            family=binomial, data = analytic.data)
require(Publish)
publish(fit0)
```

##	Variable	Units	OddsRatio	CI.95	p-value
##	RHC	No RHC	Ref		
##		RHC	2.21	[1.96;2.49]	<1e-04

[1] Adjusted regression

```
# adjust the exposure variable + demographics
fit1 <- glm(I(Death=="Yes")~RHC + age + sex + race,
            family=binomial, data = analytic.data)
publish(fit1)
```

##	Variable	Units	OddsRatio	CI.95	p-value
##	RHC	No RHC	Ref		
##		RHC	2.71	[2.38;3.08]	< 1e-04
##	age	[-Inf,50)	Ref		
##		[50,60)	3.56	[3.01;4.20]	< 1e-04
##		[60,70)	0.74	[0.64;0.87]	0.0001274
##		[70,80)	1.33	[1.15;1.54]	< 1e-04
##		[80, Inf)	1.06	[0.93;1.21]	0.3633526
##	sex	Male	Ref		
##		Female	0.49	[0.43;0.56]	< 1e-04
##	race	white	Ref		
##		black	1.10	[0.93;1.31]	0.2666157

[1] Why adjust?

Continuous outcome (Y)

- ▶ treated group $A = 1$ (RHC)
- ▶ control group $A = 0$ (no RHC)

Treatment effect = $E[Y|A = 1]$ vs. $E[Y|A = 0]$

- ▶ Would only work if 2 groups are comparable / exchangeable / ignorable treatment assignment
- ▶ Randomization with enough sample size is one

Binary outcome (Y)

Treatment effect = $\text{prob}[Y = 1|A = 1]$ vs. $\text{prob}[Y = 1|A = 0]$

[1] Why adjust?

In absence of randomization,

$$E[Y|A = 1] - E[Y|A = 0]$$

includes

- ▶ Treatment effect
- ▶ Systematic differences in 2 groups ('confounding')
 - ▶ Doctors may prescribe tx more to frail and older age patients.
 - ▶ In here, $L = \text{age}$ is a confounder.

[1] Why adjust?

In absence of randomization, if age is a known issue

Causal effect for young

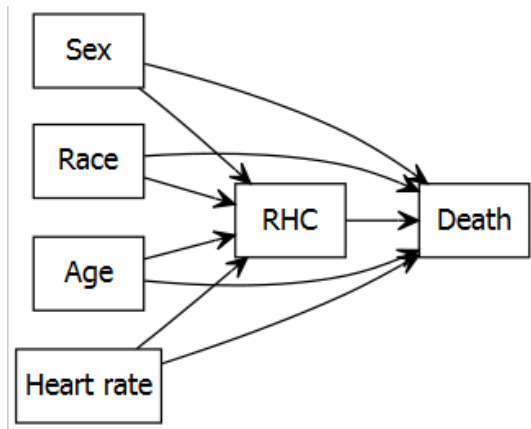
$$\blacktriangleright E[Y|A = 1, L = \text{younger age}] - E[Y|A = 0, L = \text{younger age}]$$

Causal effect for old

$$\blacktriangleright E[Y|A = 1, L = \text{older age}] - E[Y|A = 0, L = \text{older age}]$$

Conditional exchangeability; only works if L is measured

[1] Why adjust?



This was not a completely randomized data; some observational data was combined.

[1] Adjusted regression (v2)

```
# adjust the exposure variable + adjustment variables
baselinevars <- c("age", "sex", "race", "Disease.category",
                  "DNR.status", "APACHE.III.score",
                  "Pr.2mo.survival", "No.of.comorbidity",
                  "DASI.2wk.prior", "Temperature",
                  "Heart.rate", "Blood.pressure",
                  "Respiratory.rate", "WBC.count",
                  "PaO2.by.FIO2", "PaCO2", "pH",
                  "Creatinine", "Albumin", "GComa.Score")
out.formula <- as.formula(paste("I(Death=='Yes')", "~",
                               paste(baselinevars,
                                     collapse = "+")))
out.formula
```

```
## I(Death == "Yes") ~ age + sex + race + Disease.category + DNR.status +
##     APACHE.III.score + Pr.2mo.survival + No.of.comorbidity +
##     DASI.2wk.prior + Temperature + Heart.rate + Blood.pressure +
```

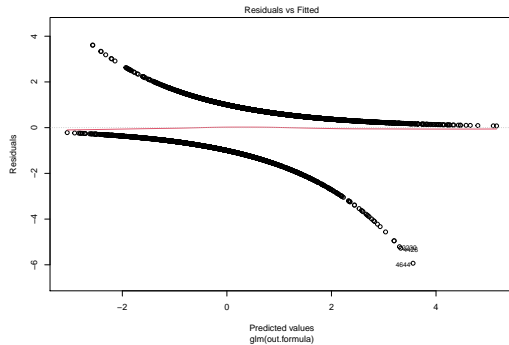

[1] Adjusted regression (v2)

```
fit2 <- glm(out.formula,  
            family=binomial, data = analytic.data)  
publish(fit2)
```

##	Variable	Units	OddsRatio	CI.95	p-value
##	age	[-Inf,50)	Ref		
##		[50,60)	1.54	[1.27;1.87]	< 1e-04
##		[60,70)	0.73	[0.62;0.87]	0.0002517
##		[70,80)	1.13	[0.97;1.32]	0.1142952
##		[80, Inf)	1.12	[0.97;1.29]	0.1117607
##	sex	Male	Ref		
##		Female	0.48	[0.42;0.55]	< 1e-04
##	race	white	Ref		
##		black	1.12	[0.93;1.36]	0.2314302
##		other	1.00	[0.74;1.35]	0.9881736
##	Disease.category	ARF	Ref		
##		CHF	1.64	[1.24;2.16]	0.0004557

[1] Adjusted regression (v2)

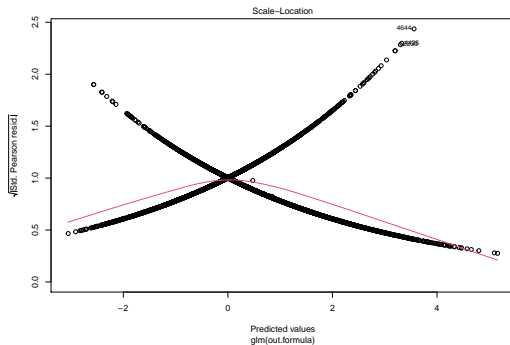
```
plot(fit2, which = 1)
```



- ▶ curvilinear trends?
 - ▶ logistic regression IS curvilinear by nature

[1] Adjusted regression (v2)

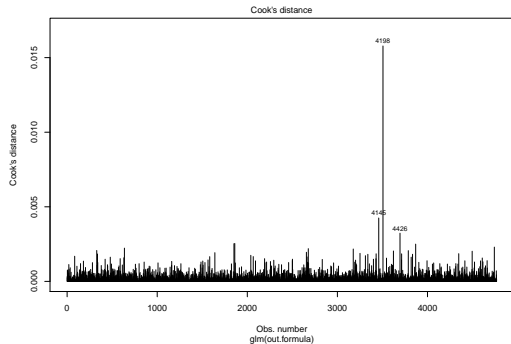
```
plot(fit2, which = 3)
```



- ▶ red line is approximately horizontal?
- ▶ points have approximately equal spread around the red line?
 - ▶ more about detecting heteroscedasticity?

[1] Adjusted regression (v2)

```
plot(fit2, which =4)
```



- Cook's D estimates the influence of data points

[2] Alternate to Regression

How sure are you about the model-specification?

- ▶ Interaction?
- ▶ Polynomial?
- ▶ Potential solution?
 - ▶ Exact Matching

[2] Exact Matching: 2 variables

```
var.comb <- do.call('paste0',  
                    analytic.data[, c('race', 'sex')])  
length(table(var.comb))
```

```
## [1] 6
```

```
table(var.comb)
```

```
## var.comb
```

```
## blackFemale  blackMale otherFemale  otherMale whiteFemale  whiteMale  
##           331         404         113         140         1340         2439
```

```
table(analytic.data$RHC,var.comb)
```

```
##      var.comb
```

```
##      blackFemale blackMale otherFemale otherMale whiteFemale whiteMale
```

```
## No RHC         161        239         50         61         667        14
```

```
## RHC           170        165         22         70         272        14
```

[2] Exact Matching: 2 variables

```
require(MatchIt)
# exact match by sex and race
m.out = matchit (RHC=="RHC" ~ sex + race,
                 data = analytic.data,
                 method = "exact")
m.out$nn
```

##	Control	Treated
## All	2583	2184
## Matched	2583	2184
## Unmatched	0	0
## Discarded	0	0

[2] Exact Matching: 3 variables

```
var.comb <- do.call('paste0',  
                    analytic.data[, c('race', 'sex', 'age')])  
length(table(var.comb))
```

```
## [1] 30
```

```
table(analytic.data$RHC, var.comb=="otherMale[80, Inf)")
```

```
##  
##          FALSE TRUE  
## No RHC    2580    3  
## RHC       2183    1
```

```
table(analytic.data$RHC, var.comb=="otherFemale[80, Inf)")
```

```
##  
##          FALSE TRUE
```


[2] Exact Matching: 3 variables

```
# exact match by age, sex and race
m.out = matchit (RHC=="RHC" ~ age + sex + race,
                 data = analytic.data,
                 method = "exact")
m.out$nn
```

##	Control	Treated
## All	2583	2184
## Matched	2581	2184
## Unmatched	0	0
## Discarded	0	0

[2] Exact Matching: 3 variables

```
matched.data <- match.data(m.out)
dim(matched.data)
```

```
## [1] 4765    25
```

```
nrow(analytic.data)-nrow(matched.data) # subjects deleted
```

```
## [1] 2
```

```
# Not taking into account of matched sets
```

```
fit1m <- glm(I(Death=="Yes")~RHC,
             family=binomial, data = matched.data)
publish(fit1m)
```

```
## Variable Units OddsRatio      CI.95 p-value
##          RHC No RHC      Ref
##          RHC      2.21 [1.96;2.49] <1e-04
```

[2] Exact Matching: many categorical variables

```
m.out = matchit (RHC=="RHC" ~ age + sex + race +  
                  Disease.category + DNR.status,  
                  data = analytic.data,  
                  method = "exact")  
  
m.out$nn
```

##	Control	Treated
## All	2583	2184
## Matched	2524	2150
## Unmatched	0	0
## Discarded	0	0

[2] Exact Matching: many categorical variables

```
matched.data <- match.data(m.out)
dim(matched.data)
```

```
## [1] 4674 25
```

```
fit2m <- glm(I(Death=="Yes")~RHC,
             family=binomial, data = matched.data)
publish(fit2m)
```

```
## Variable Units OddsRatio      CI.95 p-value
##      RHC No RHC      Ref
##           RHC      2.23 [1.98;2.51] <1e-04
```

[2] Exact Matching: including a continuous variable

```
m.out = matchit (RHC=="RHC" ~ age + sex + race +  
                 Disease.category + DNR.status +  
                 Heart.rate, # continuous  
                 data = analytic.data,  
                 method = "exact")  
  
m.out$nn
```

##	Control	Treated
## All	2583	2184
## Matched	929	947
## Unmatched	0	0
## Discarded	0	0

[2] Exact Matching: including more continuous variables

```
m.out = matchit (RHC=="RHC" ~ age + sex + race +  
                Disease.category + DNR.status +  
                Heart.rate + Blood.pressure +  
                Temperature,  
                data = analytic.data,  
                method = "exact")  
  
m.out$nn
```

##	Control	Treated
## All	2583	2184
## Matched	3	3
## Unmatched	0	0
## Discarded	0	0

[2] Exact Matching: including more continuous variables

```
matched.data <- match.data(m.out)
dim(matched.data)
```

```
## [1] 6 25
```

```
nrow(analytic.data)-nrow(matched.data) # subjects deleted
```

```
## [1] 4761
```

```
fit3m <- glm(I(Death=="Yes")~RHC,
             family=binomial, data = matched.data)
publish(fit3m)
```

```
## Variable Units OddsRatio          CI.95 p-value
##          RHC No RHC          Ref
##          RHC          1.00 [0.03;29.81]          1
```

[3] Propensity Score

Defining Propensity score (PS)

- ▶ Conditional Probability of getting treatment, given the observed covariates
- ▶ $\text{Prob}(\text{treatment: } A = 1 \mid \text{baseline or pre-treatment covariates: } L)$
 - ▶ $\text{Prob}(\text{RHC} = \text{treated} / \text{RHC group} \mid \text{age, sex, race, etc.})$
 - ▶ $f(L) = \text{Prob}(A=1|L)$

```
baselinevars
```

```
## [1] "age" "sex" "race"
## [4] "Disease.category" "DNR.status" "APACHE.III.score"
## [7] "Pr.2mo.survival" "No.of.comorbidity" "DASI.2wk.prior"
## [10] "Temperature" "Heart.rate" "Blood.pressure"
## [13] "Respiratory.rate" "WBC.count" "PaO2.by.FIO2"
## [16] "PaCO2" "pH" "Creatinine"
## [19] "Albumin" "GComa.Score"
```


[3] Propensity Score

The central role of the propensity score in observational studies for causal effects

[PR Rosenbaum](#), [DB Rubin](#) - *Biometrika*, 1983 - [academic.oup.com](#)

The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates. Both large and small sample theory show that adjustment for the scalar propensity score is sufficient to remove bias due to all observed covariates. Applications include: (i) matched sampling on the univariate propensity score, which is a generalization of discriminant matching, (ii) multivariate adjustment by subclassification on the propensity score where the same subclasses are used to estimate ...

☆ 99 Cited by 26980 Related articles All 24 versions Import into BibTeX

Theoretical result

Rosenbaum, Rubin (1983) showed:

- ▶ For potential outcomes (Y^0, Y^1) , if you have sufficient observed covariate list L to reduce confounding ('strong ignorability'): A being treatment assignment here:
 - ▶ i.e., if $(Y^0, Y^1) \perp\!\!\!\perp A|L$ (Note that is this NOT $Y \perp\!\!\!\perp A|L$)
- ▶ then
 - ▶ $(Y^0, Y^1) \perp\!\!\!\perp A|PS$ and
 - ▶ $A \perp\!\!\!\perp L|PS$

[3] Propensity Score

Assumptions

- ▶ no unmeasured confounding
- ▶ positivity ($0 < PS < 1$)
- ▶ well-defined treatment
- ▶ sufficient overlap
- ▶ model-specification

[3] Propensity Score

Variable selection for propensity score models

MA Brookhart, S Schneeweiss... - American journal of ..., 2006 - academic.oup.com

Despite the growing popularity of **propensity score** (PS) methods in epidemiology, relatively little has been written in the epidemiologic literature about the problem of **variable selection** for PS models. The authors present the results of two simulation studies designed to help ...

☆ ⓘ Cited by 1550 Related articles All 18 versions Import into BibTeX

The design *versus* the analysis of observational studies for causal effects: parallels with the design of randomized trials

DB Rubin - Statistics in medicine, 2007 - Wiley Online Library

For estimating causal effects of treatments, randomized experiments are generally considered the gold standard. Nevertheless, they are often infeasible to conduct for a variety of reasons, such as ethical concerns, excessive expense, or timeliness. Consequently, much of our knowledge of causal effects must come from non-randomized observational studies. This article will advocate the position that observational studies can and should be designed to approximate randomized experiments as closely as possible. In particular, observational ...

☆ ⓘ Cited by 896 Related articles All 6 versions Import into BibTeX

- ▶ Observed covariates are used to fix design
- ▶ Which covariates should be selected:
 - ▶ known to be a confounder (causes of Death and RHC)
 - ▶ known to be a cause of the outcome (risk factors of Death)
 - ▶ avoid known instruments or noise variables: **SE suffers**
 - ▶ mediating factors should be avoided (total effect = goal)
- ▶ Stepwise (p-value or criterion based) not recommended
 - ▶ depending on sample size, different values can get selected
 - ▶ may select variables highly associated with A

[3] Propensity Score

Many ways to use propensity scores (PS) in the analysis

- ▶ **PS matching** [our focus today]
- ▶ PS weighting
- ▶ PS stratification
- ▶ PS used as a covariate

[3] Propensity Score Matching

A tutorial and case study in **propensity score** analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality

[PC Austin](#) - Multivariate behavioral research, 2011 - Taylor & Francis

Propensity score methods allow investigators to estimate causal treatment effects using observational or nonrandomized data. In this article we provide a practical illustration of the appropriate steps in conducting propensity score analyses. For illustrative purposes, we use ...

☆ ⓘ Cited by 304 Related articles All 13 versions Import into BibTeX

Propensity score matching has 4 steps

- ▶ Stage 1: exposure modelling: $PS = Prob(A = 1|L)$
- ▶ Stage 2: Match by PS
- ▶ Stage 2: Assess balance and overlap (PS and L)
- ▶ phase 4: outcome modelling: $Prob(Y = 1|A = 1)$

[3] Propensity Score Matching

An introduction to **propensity score** methods for reducing the effects of confounding in observational studies

[PC Austin](#) - Multivariate behavioral research, 2011 - Taylor & Francis

The propensity score is the probability of treatment assignment conditional on observed baseline characteristics. The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics ...

☆ ⓘ Cited by 5235 Related articles All 15 versions Import into BibTeX

- ▶ Assessment of Balance in the whole data
 - ▶ balance = similarity of the covariate distributions
 - ▶ d or $SMD > 0.1$ can be considered as imbalance

$$d = \frac{(\bar{x}_{treatment} - \bar{x}_{control})}{\sqrt{\frac{s_{treatment}^2 + s_{control}^2}{2}}}$$

$$d = \frac{(\hat{p}_{treatment} - \hat{p}_{control})}{\sqrt{\frac{\hat{p}_{treatment}(1 - \hat{p}_{treatment}) + \hat{p}_{control}(1 - \hat{p}_{control})}{2}}}$$

```
table <- CreateTableOne(vars = baselinevars,  
                        data = analytic.data, strata = "RHC",  
                        includeNA = TRUE,
```

[3] Propensity Score Matching

```
print(table, smd = TRUE)
```

##		Stratified by RHC		
##		No RHC	RHC	SMD
##	n	2583	2184	
##	age (%)			0.181
##	[-Inf,50)	573 (22.2)	540 (24.7)	
##	[50,60)	432 (16.7)	371 (17.0)	
##	[60,70)	638 (24.7)	577 (26.4)	
##	[70,80)	603 (23.3)	529 (24.2)	
##	[80, Inf)	337 (13.0)	167 (7.6)	
##	sex = Female (%)	878 (34.0)	906 (41.5)	0.155
##	race (%)			0.098
##	white	2072 (80.2)	1707 (78.2)	
##	black	400 (15.5)	335 (15.3)	
##	other	111 (4.3)	142 (6.5)	
##	Disease category (%)			0.557

[3] Propensity Score Matching

Step 1: PS estimation

Specify the propensity score model to estimate propensity scores, and fit the model

```
ps.formula <- as.formula(paste("I(RHC == 'RHC')", "~",  
                               paste(baselinevars, collapse = "+")))
```

```
ps.formula
```

```
## I(RHC == "RHC") ~ age + sex + race + Disease.category + DNR.status +  
##   APACHE.III.score + Pr.2mo.survival + No.of.comorbidity +  
##   DASI.2wk.prior + Temperature + Heart.rate + Blood.pressure +  
##   Respiratory.rate + WBC.count + PaO2.by.FIO2 + PaCO2 + pH +  
##   Creatinine + Albumin + GComa.Score
```

- ▶ Coef of PS model fit is not of concern
- ▶ Model can be rich: to the extent that prediction is better
- ▶ But look for multi-collinearity issues
 - ▶ SE too high?

[3] Propensity score Matching

While PS has balancing property, PS is unknown and needs to be estimated:

```
# fit logistic regression to estimate propensity scores
PS.fit <- glm(ps.formula,family="binomial",
              data=analytic.data)
# extract estimated propensity scores from the fit
analytic.data$PS <- predict(PS.fit,
                           newdata = analytic.data, type="response")
```

- ▶ Other machine learning alternatives are possible to use instead of logistic regression.
 - ▶ tree based methods have better ability to detect non-linearity / non-additivity (model-specification aspect)
 - ▶ shrinkage methods - lasso / elastic net may better deal with multi-collinearity
 - ▶ ensemble learners / super learners were successfully used
 - ▶ shallow/deep learning!

[3] Propensity score Matching

- ▶ Don't lose sight that better *balance* is the ultimate goal for propensity score
- ▶ Prediction of A is just a means to that end (as true PS is unknown).
- ▶ May attract variables highly associated with A

Improving propensity score estimators' robustness to model misspecification using super learner

[R Pirracchio, ML Petersen](#)... - American journal of ..., 2015 - academic.oup.com

The consistency of propensity score (PS) estimators relies on correct specification of the PS model. The PS is frequently estimated using main-effects logistic regression. However, the underlying model assumptions may not hold. Machine learning methods provide an alternative nonparametric approach to PS estimation. In this simulation study, we evaluated the benefit of using Super Learner (SL) for PS estimation. We created 1,000 simulated data sets ($n=500$) under 4 different scenarios characterized by various degrees of deviance from ...

☆ ⓘ Cited by 98 Related articles All 9 versions Import into BibTeX

Should a propensity score model be super? The utility of ensemble procedures for causal adjustment

[S Alam, EEM Moodie, DA Stephens](#) - Statistics in medicine, 2019 - Wiley Online Library

In investigations of the effect of treatment on outcome, the propensity score is a tool to eliminate imbalance in the distribution of confounding variables between treatment groups. Recent work has suggested that Super Learner, an ensemble method, outperforms logistic regression in nonlinear settings; however, experience with real-data analyses tends to show overfitting of the propensity score model using this approach. We investigated a wide range of simulated settings of varying complexities including simulations based on real data to ...

☆ ⓘ Cited by 6 Related articles All 5 versions Import into BibTeX

[HTML] Propensity score estimation: machine learning and classification methods as alternatives to logistic regression

[D Westreich, J Lessler, MJ Funk](#) - Journal of clinical epidemiology, 2010 - ncbi.nlm.nih.gov

Objective Propensity scores for the analysis of observational data are typically estimated using logistic regression. Our objective in this Review was to assess machine learning alternatives to logistic regression which may accomplish the same goals but with fewer assumptions or greater accuracy.

☆ ⓘ Cited by 47 Related articles All 4 versions Import into BibTeX

Can we train machine learning methods to outperform the high-dimensional

[3] Propensity score Matching

Step 1

```
# summarize propensity scores  
summary(analytic.data$PS)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.  
## 0.009182 0.268112 0.454924 0.458150 0.640362 0.975476
```

```
# summarize propensity scores by exposure group  
tapply(analytic.data$PS, analytic.data$RHC, summary)
```

```
## $`No RHC`  
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.  
## 0.009182 0.184909 0.330687 0.357838 0.504012 0.974095  
##  
## $RHC  
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
```

[3] Propensity Score Matching

Step 2: PS matching

```
logitPS <- -log(1/analytic.data$PS - 1)
# logit of the propensity score
.2*sd(logitPS) # suggested in the literature
```

```
## [1] 0.2382708
```

```
0.1*sd(logitPS) # we are using this
```

```
## [1] 0.1191354
```

choosing too strict PS has unintended consequences

[PDF] [Why propensity scores should not be used for matching](#)

G King, R Nielsen - Copy at <http://j.mp/1sexgVw> Download ..., 2016 - gking.harvard.edu

We show that propensity score matching (PSM), an enormously popular method of preprocessing data for causal inference, often accomplishes the opposite of its intended goal—thus increasing imbalance, inefficiency, model dependence, and bias. The weakness ...

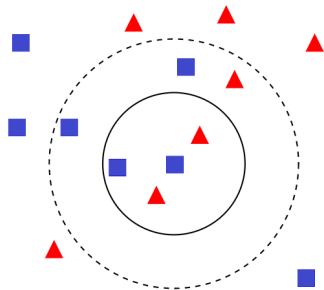
☆ ⓘ Cited by 620 Related articles All 12 versions Import into BibTeX ⓘ

[3] Propensity Score Matching

Step 2

Match using estimates propensity scores

- ▶ nearest-neighbor (NN) matching
- ▶ without replacement
- ▶ with caliper = $.1 * \text{SD of logit of propensity score}$
- ▶ with 1:1 ratio (pair-matching)



[3] Propensity Score Matching

Step 2

Match using estimated propensity scores

```
set.seed(123)
match.obj <- matchit(ps.formula, data = analytic.data,
                     distance = analytic.data$PS,
                     method = "nearest", replace=FALSE,
                     caliper = .1*sd(logitPS), ratio = 1)
# see matchit function options here
# https://www.rdocumentation.org/packages/MatchIt/versions/1.0-1/topics/ma
analytic.data$PS <- match.obj$distance
summary(match.obj$distance)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.009182 0.268112 0.454924 0.458150 0.640362 0.975476
```

[3] Propensity Score Matching

Step 2

```
match.obj$nn
```

##	Control	Treated
## All	2583	2184
## Matched	1519	1519
## Unmatched	1064	665
## Discarded	0	0

[3] Propensity Score Matching

Step 2

Step 1 and 2 can be done together by specifying distance

```
match.obj <- matchit(ps.formula, data = analytic.data,  
                     distance = 'logit',  
                     method = "nearest",  
                     replace=FALSE,  
                     caliper = .1*sd(logitPS),  
                     ratio = 1)  
analytic.data$PS <- match.obj$distance  
summary(match.obj$distance)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.  
## 0.009182 0.268112 0.454924 0.458150 0.640362 0.975476
```


[3] Propensity Score Matching

Step 2: Taking a closer look at the matches

Ref: <https://lists.gking.harvard.edu/pipermail/matchit/2013-October/0005>

```
matches <- as.data.frame(match.obj$match.matrix)
colnames(matches)<-c("matched_unit")
matches$matched_unit<-as.numeric(
  as.character(matches$matched_unit))
matches$treated_unit<-as.numeric(rownames(matches))
matches.only<-matches[!is.na(matches$matched_unit),]
head(matches.only)
```

##	matched_unit	treated_unit
## 5	438	5
## 10	2385	10
## 12	4177	12
## 13	4429	13
## 17	5228	17

[3] Propensity Score Matching

Step 2: Taking a closer look at the matches (1st pair)

```
analytic.data[analytic.data$ID %in%  
               as.numeric(matches.only[1,]),]
```

##	age	sex	race	Disease.category	DNR.status	APACHE.III.score	
## 5	[60,70)	Male	white	MOSF	Yes	72	
## 438	[80, Inf)	Male	white	ARF	Yes	93	
##	Pr.2mo.survival	No.of.comorbidity	DASI.2wk.prior	Temperature	Heart.		
## 5	0.43699980	0	21.05078	34.79688			
## 438	0.01399994	2	15.95312	34.89844			
##	Blood.pressure	Respiratory.rate	WBC.count	PaO2.by.FIO2	PaCO2		
## 5	65	27	29.699219	478.0000	17 7.2294		
## 438	46	0	8.699219	138.0938	82 7.0195		
##	Creatinine	Albumin	GComa.Score	RHC	Death	ID	PS
## 5	3.599609	3.500000	41	RHC	Yes	5	0.4213309
## 438	3.500000	2.799805	100	No BHC	Yes	438	0.4125774

[3] Propensity Score Matching

Step 2: Taking a closer look at the matches (2nd pair)

```
analytic.data[analytic.data$ID %in%  
               as.numeric(matches.only[2,]),]
```

##	age	sex	race	Disease.category	DNR.status	APACHE.III.score	
## 10	[-Inf,50)	Female	white	ARF	No	4	
## 2385	[70,80)	Female	white	ARF	No	3	
##	Pr.2mo.survival	No.of.comorbidity	DASI.2wk.prior	Temperature	Heart		
## 10	0.6689453	1	23.25781	38.5			
## 2385	0.6219997	1	18.35156	36.5			
##	Blood.pressure	Respiratory.rate	WBC.count	PaO2.by.FIO2	PaCO2		
## 10	73	40	20.597656	68.0000	30 7.3490		
## 2385	67	9	3.199707	168.5625	34 7.4290		
##	Creatinine	Albumin	GComa.Score	RHC	Death	ID	PS
## 10	0.500000	2.5	0	RHC	No	10	0.6565211
## 2385	1.599854	3.5	0	No BHC	No	2385	0.6292985

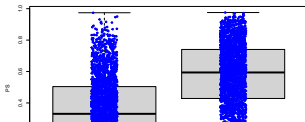
[3] Propensity score Matching

Step 3: Assessing balance and overlap

Balance is more important than prediction!

- ▶ Criteria to assess success of step 2: PS estimation
 - ▶ better balance
 - ▶ better overlap
 - ▶ $PS = 0$ or $PS = 1$ needs close inspection

```
boxplot(PS ~ RHC=='RHC', data = analytic.data,  
        lwd = 2, ylab = 'PS')  
stripchart(PS ~ RHC=='RHC', vertical = TRUE,  
           data = analytic.data, method = "jitter",  
           add = TRUE, pch = 20, col = 'blue')
```

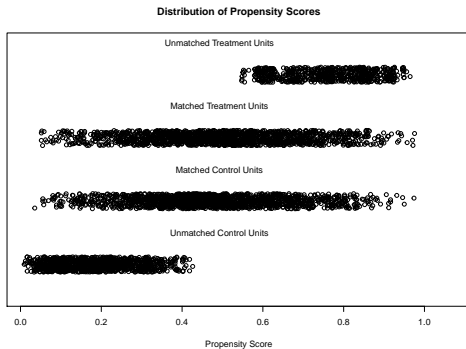


[3] Propensity Score Matching

Step 3

Vizualization

```
plot(match.obj, type = "jitter")
```



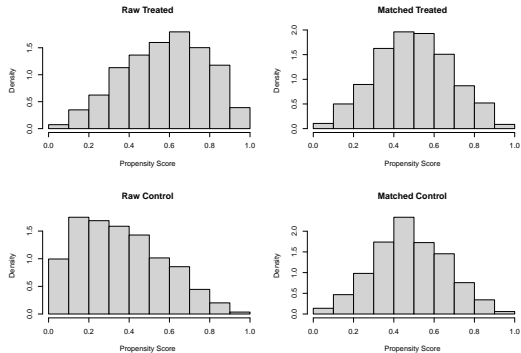
```
## [1] "To identify the units, use first mouse button; to stop, use second
```

[3] Propensity Score Matching

Step 3

Vizualization for assessing overlap issues

```
plot(match.obj, type = "hist")
```



[3] Propensity Score Matching

Step 3

Assessment of Balance

```
matched.data <- match.data(match.obj)
tab1m <- CreateTableOne(vars = baselinevars,
                        data = matched.data, strata = "RHC",
                        includeNA = TRUE,
                        test = TRUE, smd = TRUE)
```

Compare the similarity of baseline characteristics between treated and untreated subjects in a the propensity score-matched sample.

- ▶ In this case, we will compare $SMD < 0.1$ or not.
- ▶ In some literature, other generous values (0.25) are proposed.

An introduction to **propensity score** methods for reducing the effects of confounding in observational studies

[PC Austin](#) - Multivariate behavioral research, 2011 - Taylor & Francis

The propensity score is the probability of treatment assignment conditional on observed baseline characteristics. The propensity score allows one to design and analyze an observational (nonrandomized) study so that it mimics some of the particular characteristics ...

☆ ⓘ Cited by 5235 Related articles All 15 versions Import into BibTeX

[3] Propensity Score Matching

Step 3

```
print(tab1m, showAllLevels = FALSE, smd = TRUE, test = FALSE)
```

		Stratified by RHC		
		No RHC	RHC	SMD
##	n	1519	1519	
##	age (%)			0.042
##	[-Inf,50)	367 (24.2)	365 (24.0)	
##	[50,60)	255 (16.8)	272 (17.9)	
##	[60,70)	385 (25.3)	395 (26.0)	
##	[70,80)	368 (24.2)	351 (23.1)	
##	[80, Inf)	144 (9.5)	136 (9.0)	
##	sex = Female (%)	560 (36.9)	572 (37.7)	0.016
##	race (%)			0.017
##	white	1212 (79.8)	1212 (79.8)	
##	black	229 (15.1)	234 (15.4)	

[3] Propensity Score Matching

Step 3

Possible to get p-values to check balance: but strongly discouraged

- P-value based balance assessment can be influenced by sample size

A tutorial and case study in **propensity score** analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality

[PC Austin](#) - Multivariate behavioral research, 2011 - Taylor & Francis

Propensity score methods allow investigators to estimate causal treatment effects using observational or nonrandomized data. In this article we provide a practical illustration of the appropriate steps in conducting propensity score analyses. For illustrative purposes, we use ...

☆ 99 Cited by 304 Related articles All 13 versions Import into BibTeX

```
print(tab1m, showAllLevels = FALSE, smd = FALSE, test = TRUE)
```

##		Stratified by RHC		
##		No RHC	RHC	p
##	n	1519	1519	
##	age (%)			0.859
##	[-Inf,50)	367 (24.2)	365 (24.0)	
##	[50,60)	255 (16.8)	272 (17.9)	
##	[60,70)	385 (25.3)	395 (26.0)	

[3] Propensity Score Matching

Step 3

Assessment of balance in the matched data

```
smd.res <- ExtractSmd(tab1m)
t(round(smd.res,2))
```

```
##          age  sex race Disease.category DNR.status APACHE.III.score
## 1 vs 2 0.04 0.02 0.02              0.04          0.01              0.07
##          Pr.2mo.survival No.of.comorbidity DASI.2wk.prior Temperature Hear
## 1 vs 2          0.06              0.01          0          0.01
##          Blood.pressure Respiratory.rate WBC.count PaO2.by.FIO2 PaCO2    pl
## 1 vs 2          0.05              0.04          0.01          0.04  0.01  0.0
##          Creatinine Albumin GComa.Score
## 1 vs 2          0.02          0.03          0.01
```

[3] Propensity Score Matching

Step 3: Variance ratio

- ▶ Variance ratios ~ 1 means:
- ▶ equal variances in groups
- ▶ group balance
- ▶ could vary from 1/2 to 2
- ▶ other cut-points are suggested as well (0.8 to 1.2)

Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples

[PC Austin](#) - Statistics in medicine, 2009 - Wiley Online Library

The propensity score is a subject's probability of treatment, conditional on observed baseline covariates. Conditional on the true propensity score, treated and untreated subjects have similar distributions of observed baseline covariates. Propensity-score matching is a popular method of using the propensity score in the medical literature. Using this approach, matched sets of treated and untreated subjects with similar values of the propensity score are formed. Inferences about treatment effect made using propensity-score matching are ...

☆ ⓘ Cited by 2214 Related articles All 11 versions Import into BibTeX

[HTML] Matching methods for causal inference: A review and a look forward

[EA Stuart](#) - Statistical science: a review journal of the Institute of ..., 2010 - ncbi.nlm.nih.gov

When estimating causal effects using observational data, it is desirable to replicate a randomized experiment as closely as possible by obtaining treated and control groups with similar covariate distributions. This goal can often be achieved by choosing well-matched samples of the original treated and control groups, thereby reducing bias due to the covariates. Since the 1970's, work on matching methods has examined how to best choose treated and control subjects for comparison. Matching methods are gaining popularity in ...

☆ ⓘ Cited by 3066 Related articles All 22 versions Import into BibTeX ⓘ

[3] Propensity Score Matching

Step 3: Variance ratio

```
require(cobalt)
baltab.res <- bal.tab(x = match.obj, data = analytic.data,
                     treat = analytic.data$RHC,
                     disp.v.ratio = TRUE)
```

```
## Note: 's.d.denom' not specified; assuming pooled.
```

```
baltab.res$Balance$V.Ratio.Adj
```

```
## [1] 1.0990553      NA      NA      NA      NA      NA
## [8]      NA      NA      NA      NA      NA      NA
## [15]      NA 1.0867497 0.9714495 0.9605864 0.8305596 0.8395535 0.94083
## [22] 0.9841995 1.0655834 1.0262382 0.9733399 1.0919443 1.0916685 0.61000
## [29] 2.0325397 0.9847091
```

[3] Propensity Score Matching

Step 4: Outcome modelling

- ▶ Some flexibility in choosing outcome model
 - ▶ considered independent of exposure modelling
 - ▶ some propose double robust approach
 - ▶ adjusting imbalanced covariates only?

Estimate the effect of treatment on outcomes using propensity score-matched sample

```
fit3 <- glm(I(Death=="Yes")~RHC,  
            family=binomial, data = matched.data)  
publish(fit3)
```

##	Variable	Units	OddsRatio	CI.95	p-value
##	RHC	No RHC	Ref		
##		RHC	2.08	[1.80;2.41]	<1e-04

[3] Propensity Score Matching

Step 4: Outcome modelling

```
out.formula
```

```
## I(Death == "Yes") ~ age + sex + race + Disease.category + DNR.status +  
##     APACHE.III.score + Pr.2mo.survival + No.of.comorbidity +  
##     DASI.2wk.prior + Temperature + Heart.rate + Blood.pressure +  
##     Respiratory.rate + WBC.count + PaO2.by.FIO2 + PaCO2 + pH +  
##     Creatinine + Albumin + GComa.Score
```

```
fit3b <- glm(out.formula,  
             family=binomial, data = matched.data)  
publish(fit3b)
```

##	Variable	Units	OddsRatio	CI.95	p-value
##	age	[-Inf,50)	Ref		
##		[50,60)	1.62	[1.26;2.09]	0.0001639

[3] Propensity Score Matching

Step 4: Other considerations for outcome model

The above analysis do not take matched pair into consideration while regressing. Literature proposes different strategies:

- ▶ do not control for pairs / clusters
 - ▶ use `glm` as is
- ▶ control for pairs / clusters
 - ▶ use `cluster` option or GEE or conditional logistic
- ▶ Bootstrap for matched pairfor WOR
 - ▶ may not be appropriate for WR

The use of bootstrapping when using propensity-score matching without replacement: a simulation study

[PC Austin, DS Small - Statistics in medicine, 2014 - Wiley Online Library](#)

Propensity-score matching is frequently used to estimate the effect of treatments, exposures, and interventions when using observational data. An important issue when using propensity-score matching is how to estimate the standard error of the estimated treatment effect. Accurate variance estimation permits construction of confidence intervals that have the advertised coverage rates and tests of statistical significance that have the correct type I error rates. There is disagreement in the literature as to how standard errors should be ...

☆ ⓘ Cited by 139 Related articles All 11 versions Import into BibTeX

[3] Propensity Score Matching

Step 4

- ▶ The example compared RHC (a treated group; target) vs No RHC (untreated).
- ▶ The corresponding treatment effect estimate is known as
 - ▶ Average Treatment Effects on the Treated (ATT)
- ▶ Other estimates from PS analysis are possible that compared the whole population
 - ▶ what if everyone treated vs. what if nobody was treated (ATE)

[3] Propensity Score Matching

Other matching algorithms

- ▶ Optimal
- ▶ genetic matching
- ▶ CEM
- ▶ variable ratio NN

[3] Propensity Score Matching

- ▶ MatchIt
- ▶ Matching

Other useful packages

- ▶ cobalt
- ▶ twang

Outdated package

- ▶ nonrandom

[4] Discipline-specific PS Systematic Reviews

- ▶ Propensity score matching most popular
 - ▶ Cardiovascular / Infective endocarditis / Intensive care
 - ▶ Critical care / anesthesiology / Sepsis / Psychology
 - ▶ Cancer / Multiple sclerosis
- ▶ Not meta-analysis; but reviews of usage of PS methods in different disciplines

[PTM13] Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement
RC Austin. *The Journal of thoracic and cardiovascular surgery*, 2007 - Elsevier
Objective I conducted a systematic review of the use of propensity score matching in the cardiovascular surgery literature. I examined the adequacy of the methodology and whether appropriate statistical methods were used. Methods I examined 60 articles published in the *Annals of Thoracic Surgery*, *European Journal of Cardio-thoracic Surgery*, *Journal of Cardiovascular Surgery*, and the *Journal of Thoracic and Cardiovascular Surgery* between January 1, 2004, and December 31, 2006. Results Thirty-one of the 60 studies did not ...
☆ 50 Cited by 482 Related articles All 9 versions Import into BioRx

Propensity scores in intensive care and anaesthesiology literature: a systematic review
E Gayat, R Pirracchio, M Resche-Rigon. *Intensive care*, 2010 - Springer
Introduction Propensity score methods have been increasingly used in the last 10 years. However, the practical use of the propensity score (PS) has been reported as heterogeneous in several papers reviewing the use of propensity scores and giving some advice. No precedent work has focused on the specific application of PS in intensive care and anaesthesiology literature. Objectives After a brief development of the theory of propensity score, to assess the use and the quality of reporting of PS studies in intensive ...
☆ 50 Cited by 107 Related articles All 17 versions Import into BioRx

[PTM13] Observational studies using propensity score analysis underestimated the effect sizes in critical care medicine
Z Zhang, H Ni, X Xu - *Journal of clinical epidemiology*, 2014 - Elsevier
Abstract Background and Objective Propensity score (PS) analysis has been increasingly used in critical care medicine, however, its validation has not been systematically investigated. The present study aimed to compare effect sizes in PS-based observational studies vs. randomized controlled trials (RCTs) or meta-analysis of RCTs. Methods Critical care observational studies using PS were systematically searched in PubMed from inception to April 2013. Identified PS-based studies were matched to one or more RCTs in terms of ...
☆ 50 Cited by 49 Related articles All 10 versions Import into BioRx

[PTM13] Propensity scores: methods, considerations, and applications in the Journal of Thoracic and Cardiovascular Surgery
TL McMurry, Y Hu, EH Blackstone. *The Journal of thoracic and ...*, 2015 - Elsevier
Objective To review the published literature using propensity scoring, describe shortcomings in the use of this technique, and provide conceptual background for understanding and correctly implementing studies that use propensity matching. Methods We survey the published statistical literature and make recommendations for a set of standard criteria for studies that use propensity matching. We evaluated adherence to these criteria in recent publications in the *Journal of Thoracic and Cardiovascular Surgery* and determined how ...
☆ 50 Cited by 81 Related articles All 7 versions Import into BioRx

A systematic review of propensity score methods in the social sciences
EJ Thoemmes, ES Kim - *Multivariate behavioral research*, 2011 - Taylor & Francis
The use of propensity scores in psychological and educational research has been steadily increasing in the last 2 to 3 years. However, there are some common misconceptions about the use of different estimation techniques and conditioning choices in the context of propensity score analysis. In addition, reporting practices for propensity score analyses often lack important details that allow other researchers to confidently judge the appropriateness of reported analyses and potentially to replicate published findings. In this article we conduct ...
☆ 50 Cited by 401 Related articles All 10 versions Import into BioRx

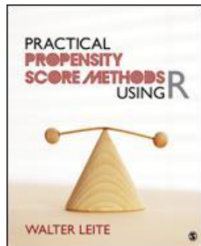
Reporting and guidelines in propensity score analysis: a systematic review of cancer and cancer surgical studies
Xi Yao, X Wang, PJ Speicher, ES Huang. *JNCI: Journal of the ...*, 2017 - academic.oup.com
Background. Propensity score (PS) analysis is increasingly being used in observational studies, especially in some cancer studies where random assignment is not feasible. This systematic review evaluates the use and reporting quality of PS analysis in oncology studies. Methods. We searched PubMed to identify the use of PS methods in cancer studies (CS) and cancer surgical studies (CSS) in major medical, cancer, and surgical journals over time and critically evaluated 33 CS published in top medical and cancer journals in 2014 ...
☆ 50 Cited by 82 Related articles All 11 versions Import into BioRx

[PTM13] The role of valve surgery in infective endocarditis management: a systematic review of observational studies that included propensity score analysis
IM Teytey, T Kashour, V Zimmerman. *American heart ...*, 2008 - Elsevier
Background The potential role of valve surgery in infective endocarditis (IE) management is controversial. No randomized trials have been conducted to date; accordingly, some studies use propensity score analysis (PSA) to minimize selection bias in observational studies. Methods A systematic review of the literature addressing the role of valve surgery in IE was performed. Studies in which PSA was applied to the management of IE were identified using Medline, Web of Science, Zetoc, and Article First from inception to June 2007. Cohort studies ...
☆ 50 Cited by 58 Related articles All 9 versions Import into BioRx

[PTM13] Do the observational studies using propensity score analysis agree with randomized controlled trials in the area of sepsis?
Z Zhang, H Ni, X Xu - *Journal of critical care*, 2014 - Elsevier
Background and objectives Sepsis is a leading cause of mortality and morbidity in the intensive care unit, and many studies have been conducted aiming to improve its outcome. Randomized controlled trials (RCTs) and observational studies using propensity score (PS) method are commonly used for this purpose. However, the agreement between these two major methodological designs has never been investigated in this specific area. The present study aimed to compare the effect sizes between RCTs and PS-based studies. Methods ...
☆ 50 Cited by 20 Related articles All 10 versions Import into BioRx

[PTM13] Evaluation of propensity score used in cardiovascular research: a cross-sectional survey and guidance document
M Samuel, B Balomen, J Rouette, J Kim, RV Platt. *BMJ open*, 2020 - bmjopen.bmj.com
Background Propensity score (PS) methods are frequently used in cardiovascular clinical research. Previous evaluations revealed poor reporting of PS methods, however a comprehensive and current evaluation of PS use and reporting is lacking. The objectives of the present survey were to (1) evaluate the quality of PS methods in cardiovascular publications, (2) summarise PS methods and (3) propose key reporting elements for PS publications. Methods A PubMed search for cardiovascular PS articles published between ...
☆ 50 All 4 versions Import into BioRx

Further Reading



[BOOK] **Practical propensity score methods using R**

[W Leite](#) - 2016 - [books.google.com](#)

Practical Propensity Score Methods Using R by Walter Leite is a practical book that uses a step-by-step analysis of realistic examples to help students understand the theory and code for implementing propensity score analysis with the R statistical language. With a ...

☆ [Cited by 67](#) [Related articles](#) [Import into BibTeX](#)

Companion site: study.sagepub.com/leite

Thank you!

ehsank.com