Applied example of Causal Inference Analysis

Brice Batomen

AI4PH Summer Institute

2022/21/11 (updated: 2023-07-13)

Outline

1. The course dataset

2. Propensity Scores

3. Inverse Probability of Treatment Weighting

4. Matching

5. Closing remarks

WORKING EXAMPLE

Concepts in Emergency and Critical Care

The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients

Alfred F. Connors, Jr, MD; Theodore Speroff, PhD; Neal V. Dawson, MD; Charles Thomas; Frank E. Harrell, Jr, PhD; Douglas Wagner, PhD; Norman Desbiens, MD; Lee Goldman, MD, MPH; Albert W. Wu, MD; Robert M. Califf, MD; William J. Fulkerson, Jr, MD; Humberto Vidaillet, MD; Steven Broste, MS; Paul Bellamy, MD; Joanne Lynn, MD; William A. Knaus, MD; for the SUPPORT Investigators

This dataset was used in Connors et al. (1996): The effectiveness of RHC in the initial care of critically ill patients. JAMA. The "treatment" variable swang1 is whether or not a patient received a RHC (also called the Swan-Ganz catheter) on the first day in which the patient qualified for the SUPPORT study.

Right heart catheterization dataset

IMPORT DATASET

```
library(here)
rhc <- as.data.frame(read.csv("rhc.csv", header=T)); rhc$A <- ifelse(rhc$swang1 =="No RHC", 0, 1)

#Variables with missing data
names(which(colSums(is.na(rhc))>0))
```

```
## [1] "cat2"    "dschdte" "dthdte"  "adld3p"  "urin1"
```

```
#Proportion of Missing data
p <- ((colMeans(is.na(rhc)))*100); p[which(p> 0)]
```

```
##        cat2     dschdte      dthdte      adld3p       urin1
## 79.07585004  0.01743679 35.10026155 74.90845684 52.79860506
```

```
##                   Stratified by swang1
##                    No RHC       RHC          p      test
##   n                 3551         2184
##   surv30 = 1 (%)    2463 (69.4)  1354 (62.0)  <0.001
##   surv60 = 1 (%)    2231 (62.8)  1190 (54.5)  <0.001
##   surv180 = 1 (%)   1905 (53.6)  1008 (46.2)  <0.001
##   death = Yes (%)   2236 (63.0)  1486 (68.0)  <0.001
```

Potential-outcome / counterfactual models

A causal effect is defined as a comparison between two states of the world, the “actual” and the “counterfactual”. (*Cunningham 2021*)

Let’s consider a treatment A that takes on a value of $A_i = 1$ if a particular individual *i* receives the treatment and $A_i = 0$ if it does not.

Each individual will have 2 POs:

$Y_i^{a=1}$, Potential Outcome if $A_i = 1$

$Y_i^{a=0}$, Potential Outcome if $A_i = 0$

The causal effect of treatment equals the difference between two separate states of the world:

- $\delta_i = Y_i^{a=1} - Y_i^{a=0}$, for the exact same person at the exact same moment in time.

EFFECT DEFINITION - TOTAL EFFECT

$E[Y^{a=1}] - E[Y^{a=0}] =$ **Average Total Effect (ATE)**

- **A** = Right Heart Catheterization (RHC); **Y** = Survival at a given day

To compute that *ATE* from observational data we need to satisfy some assumptions :

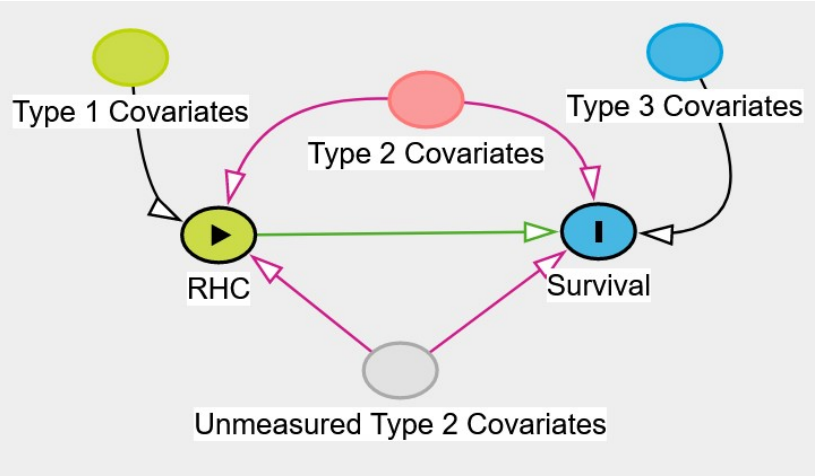
- **Consistency and no interference**
- **Conditional exchangeability**
  - No measurement errors
  - Correct model specification
- **Positivity**

If we can convince an informed audience that these assumptions hold, then we can move from the potential outcomes to observed outcomes:

- $E[Y^a] = \sum_c E[Y|A = a, C = c]Pr[C = c]$; C = Confounders.

CAUSAL FRAMEWORK

Variables	
Age	Temperature
Sex	Mean blood pressure
Race	Respiratory rate
Years of education	Heart rate
Income	PaO2/FIO2 ratio
Medical insurance (categorical)	PaCo2
Primary disease category	PH
Secondary disease category	WBC
Categories of admission diagnosis	Hematocrit
ADL 2 weeks before admission	Sodium
DASI (Duke Activity Status Index) 2 weeks before admission	Potassium
DNR status on day1 (do-not-resuscitate status)	Creatinine
Cancer (none, localized, metastatic)	Bilirubin
Support model estimate of the prob. of surviving 2 months	Albumin
APACHE score	Urine output
Glasgow Coma Score	Categories of comorbidities illness (13)
Weight	



- Subject knowledge is critical to be able to identify covariates that need to be collected to satisfy assumptions listed before
  - Ideally we want to collect all Type 2 covariates.

PROPENSITY SCORE

They are several techniques to estimate the $E[Y^a] = \sum_c E[Y|A = a, C = c]Pr[C = c]$:

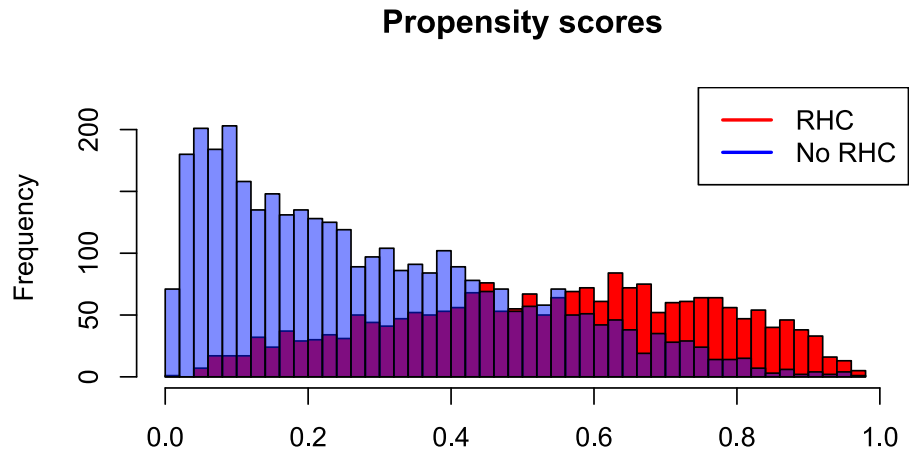
- **Marginal structural model**
  - Propensity score (PS)
- Direct Standardization (Parametric G-formula)
- Targeted Maximum Likelihood Estimation (TMLE) etc.

PS = the probability of each individual of being exposed/treated given a set of covariates (C), necessary to achieve conditional exchangeability.

- $PS = Pr[A = 1|C]$
- Usually done by fitting a model for the exposure given covariates
  - Type 2 and 3 covariates.
- In practice, the true PS $(Pr[A = 1|C])$ are not known
  - We can use ML techniques to obtain the PS.

PROPENSITY SCORE - CODE

```
d <- glm(A ~ age +factor(sex) +factor(race) +edu +factor(income) +factor(ninsclas) +factor(cat1) +resp +card
        +neuro +gastr +renal +meta +hema +seps +trauma +ortho +das2d3pc +factor(dnr1) +factor(ca) +surv2md1 +aps
        +scoma1 +wtkilo1 +temp1 +meanbp1 + resp1 +hrt1 + pafi1 +paco21 +ph1 +wblc1 +hema1 +sod1 + pot1+ crea1
        +bili1 +alb1 +factor(cardiohx) +factor(chfhx) +factor(dementhx) +factor(psychhx) +factor(chrpulhx)
        +factor(renalhx) +factor(liverhx) +factor(gibledhx) +factor(malighx) +factor(immunhx) +factor(transhx)
        +factor(amihx), data=rhc, family=binomial(link=logit))
deno <- predict(d, type="response")

hist(deno[rhc$A==1], xlim=c(0,1), ylim=c(0, 225), col="red", breaks=40, main='Propensity scores', xlab="")
hist(deno[rhc$A==0], add=T, col=rgb(0, 0.1, 1, 0.5), breaks= 40)
legend('topright', legend=c('RHC','No RHC'), lwd=2, col=c('red','blue'))
```



A. Inverse probability of treatment weighting (IPTW)

After obtaining the PS for each individual's, we need to :

- Create a weighted pseudopopulation with no-imbalances in the measured covariates
    - Among the exposed, upweight individuals who have small probability of being exposed given their covariates. If $A_i = 1$ then $IPTW_i = 1/(PS_i) = 1/(Pr[A_i = 1|C_i])$
    - Among the unexposed, upweight individuals who have small probability of being unexposed given their covariates. If $A_i = 0$ then $IPTW_i = 1/(1 - PS_i) = 1/(Pr[A_i = 0|C_i])$.
    - For precision gains, we can stabilize the weights, $SIPTW_i = (Pr[A_i = a])/(Pr[A_i = a|C_i])$.
- Assess the balance of covariates.

If the PS model is correctly specified, the resulting IPTW estimator is still asymptotically unbiased.

IPTW - code

```
n <- glm(A ~ 1, data=rhc, family=binomial(link=logit)); nume <- predict(n, type="response")

rhc$SIPTW <- (rhc$A==1) * nume/deno + (rhc$A==0) * (1-nume)/(1-deno)
summary(rhc$SIPTW)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.3894  0.6565  0.7798  0.9935  1.0554 20.0010
```

```
# Weighted covariate balance:
library(survey)
covs <- c("age", "sex", "race", "edu", "income", "ninsclas", "cat1", "resp", "card", "neuro", "gastr", "renal",

weighted <- svydesign(ids=~0, data=rhc, weights=rhc$SIPTW)
table.w <- svyCreateTableOne(vars=covs, strata="A", data=weighted, smd=TRUE, test=F)
```

Look at Table 1 of the *JAMA* paper.

Standardized mean difference (SMD) is a statistic used to examine the balance of covariate distribution between treatment groups.

SMD is not influenced by sample size and allows for the comparison of the relative balance of variables measured in different units. (*Austin & Stuart, 2015*)

IPTW Assess balance of covariates

```
# A.2 Weighted covariate balance:
print(table.w, smd=TRUE)
```

```
##                          Stratified by A
##                           0                1                SMD
##   n                       3583.85          2113.80
##   age (mean (SD))           61.17 (17.13)    61.11 (15.56)   0.004
##   sex = Male (%)          2001.6 (55.9)    1211.9 (57.3)    0.030
##   race (%)                                                  0.021
##      black                 567.4 (15.8)     348.3 (16.5)
##      other                 212.7 ( 5.9)     118.8 ( 5.6)
##      white                2803.7 (78.2)    1646.7 (77.9)
##   edu (mean (SD))          11.69 (3.12)     11.72 (3.00)    0.011
##   income (%)                                                0.026
##      $11-$25k              736.0 (20.5)     417.1 (19.7)
##      $25-$50k              538.6 (15.0)     332.1 (15.7)
##      > $50k                276.7 ( 7.7)     167.3 ( 7.9)
##      Under $11k           2032.6 (56.7)    1197.3 (56.6)
##   ninsclas (%)                                             0.042
##      Medicaid              415.8 (11.6)     253.1 (12.0)
##      Medicare              910.9 (25.4)     499.6 (23.6)
##      Medicare & Medicaid   226.0 ( 6.3)     138.8 ( 6.6)
##      No insurance          197.8 ( 5.5)     116.8 ( 5.5)
##      Private              1057.3 (29.5)     637.9 (30.2)
##      Private & Medicare    776.0 (21.7)     467.5 (22.1)
##   cat1 (%)                                                 0.078
```

IPTW - Weighting the outcomes

```
library("geepack")
library(sjPlot)
MSM <- geeglm(surv30 ~ swang1, family=binomial("log"), data=rhc,
              weights=SIPTW, std.err = 'san.se', id=ptid, corstr="independence")
MSM.1 <- geeglm(surv30 ~ swang1, family=binomial("logit"), data=rhc,
              weights=SIPTW, std.err = 'san.se', id=ptid, corstr="independence")
tab_model(MSM, MSM.1)
```

	surv 30			surv 30		
Predictors	*Risk Ratios*	*CI*	*p*	*Odds Ratios*	*CI*	*p*
(Intercept)	0.69	0.67 – 0.70	**<0.001**	2.18	2.00 – 2.38	**<0.001**
swang1 [RHC]	0.92	0.88 – 0.97	**0.002**	0.79	0.68 – 0.91	**0.002**
N	5735 ptid			5735 ptid		
Observations	5735			5735		

```
MSM.60 <- geeglm(surv60 ~ swang1, family=binomial("log"), data=rhc,
              weights=SIPTW, std.err = 'san.se', id=ptid, corstr="independence")
MSM.180 <- geeglm(surv180 ~ swang1, family=binomial("log"), data=rhc,
              weights=SIPTW, std.err = 'san.se', id=ptid, corstr="independence")
```

IPTW - Weighting the outcomes(2)

```
tab_model(MSM, MSM.60, MSM.180)
```

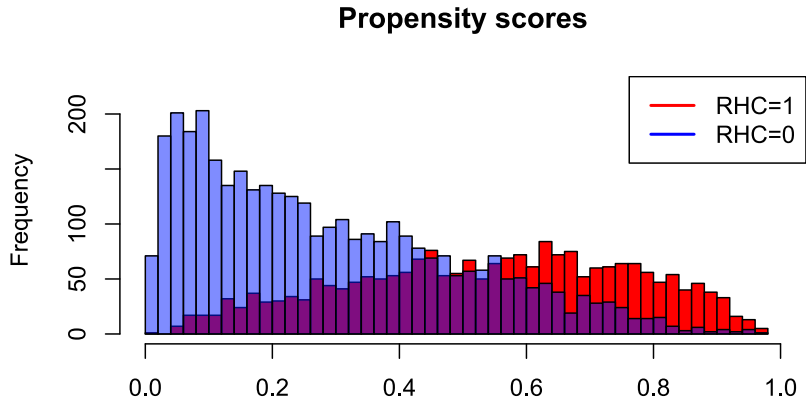
	surv 30			surv 60			surv 180		
Predictors	*Risk Ratios*	*CI*	*p*	*Risk Ratios*	*CI*	*p*	*Risk Ratios*	*CI*	*p*
(Intercept)	0.69	0.67 – 0.70	**<0.001**	0.62	0.60 – 0.64	**<0.001**	0.53	0.51 – 0.55	**<0.001**
swang1 [RHC]	0.92	0.88 – 0.97	**0.002**	0.91	0.86 – 0.97	**0.002**	0.91	0.85 – 0.98	**0.009**
N	5735 ptid			5735 ptid			5735 ptid		
Observations	5735			5735			5735		

```
A1  <- rhc; A1$swang1 <- "RHC"; A0  <- rhc; A0$swang1 <- "No RHC";
P.A1 <- predict(MSM.1, newdata=data.frame(A1), type="response")
P.A0 <- predict(MSM.1, newdata=data.frame(A0), type="response")

RD30 <- mean(P.A1) - mean(P.A0); RR.30 <- mean(P.A1) / mean(P.A0);
OR.30 <- mean(P.A1/(1-P.A1)) / mean(P.A0/(1-P.A0)); OR.30a <- mean(P.A0/(1-P.A0)) / mean(P.A1/(1-P.A1))
round(cbind(RD30, RR.30, OR.30, OR.30a), 2)
```

```
##       RD30 RR.30 OR.30 OR.30a
## [1,] -0.05  0.92  0.79   1.27
```

B. Matching

```
hist(deno[rhc$A==1], xlim=c(0,1), ylim=c(0, 225), col="red", breaks=40, main='Propensity scores', xlab="")
hist(deno[rhc$A==0], add=T, col=rgb(0, 0.1, 1, 0.5), breaks= 40)
legend('topright', legend=c('RHC=1','RHC=0'), lwd=2, col=c('red','blue'))
```



- We restrict the analyses to participants within the common support zone.
- They are several matching specifications(Exact, Nearest neighbors, Optimal etc.)
  - However, always check the balance of covariates.
- Several statistical packages exist to perform matching.

Matching and Target population

- Note that the matching procedure may distort the study sample
  - Therefore, the estimated treatment effect might not correspond to a well defined population.
- We can compute marginal effects for different populations:
  - $ATE = E[Y^1] - E[Y^0]$: average effect of the treatment for all participants in the target population
  - $ATT = E[Y^1|A = 1] - E[Y^0|A = 1]$: ~ participants like those who actually were treated
  - $ATU = E[Y^1|A = 0] - E[Y^0|A = 0]$: ~ participants like those who actually were untreated.
- Most matching procedures aim to estimate the **ATT**, but because some participants might be discarded, the result obtained is the *average treatment effect in the matched sample*.

Matching - CODES

- In the *JAMA* paper they used the nearest neighbors procedure, with exact matching on disease category (page 891 - Case-Matching Procedure).

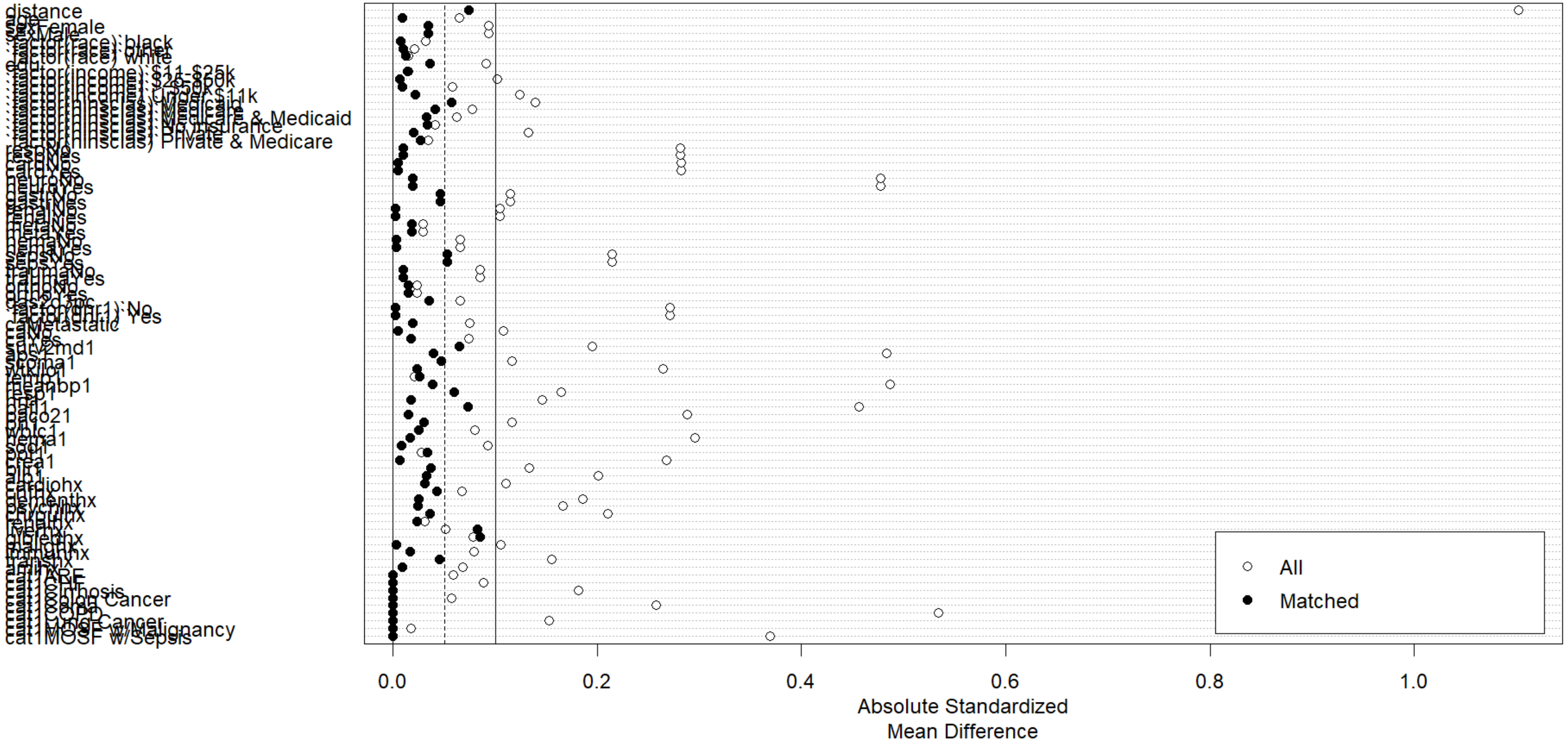
```
library(MatchIt)
formula <- A ~ age +sex +factor(race) +edu +factor(income) +factor(ninsclas) +resp +card +neuro +gastr +renal +m

m.out1 <- matchit(formula, data = rhc, method = "nearest", exact= ~ cat1, distance = "glm", caliper = .15)
```

- Check the instructions of the MatchIt package here: MatchIt

- The literature on propensity score matching is still evolving.

Matching - CODES (2)

Checking balance after nearest neighbors matching



Matching - CODES (3)

```
Match <- match.data(m.out1)

PSM.30 <- geeglm(surv30 ~ swang1, family=binomial("log"), data=Match,
              weights=weights, std.err = 'san.se', id=subclass, corstr="independence")
PSM.60 <- geeglm(surv60 ~ swang1, family=binomial("log"), data=Match,
              weights=weights, std.err = 'san.se', id=subclass, corstr="independence")
PSM.180 <- geeglm(surv180 ~ swang1, family=binomial("log"), data=Match,
              weights=weights, std.err = 'san.se', id=subclass, corstr="independence")
tab_model(PSM.30, PSM.60, PSM.180)
```

	surv 30			surv 60			surv 180		
Predictors	*Risk Ratios*	*CI*	*p*	*Risk Ratios*	*CI*	*p*	*Risk Ratios*	*CI*	*p*
(Intercept)	0.70	0.68 – 0.72	**<0.001**	0.63	0.61 – 0.65	**<0.001**	0.53	0.51 – 0.56	**<0.001**
swang1 [RHC]	0.90	0.85 – 0.94	**<0.001**	0.88	0.83 – 0.93	**<0.001**	0.87	0.81 – 0.93	**<0.001**
N	1597 subclass			1597 subclass			1597 subclass		
Observations	3194			3194			3194		

C. OTHER APPROACHES WITH PS

- Subclassification
  - Groups individuals into subclasses based on their propensity score values
  - Effect estimates are obtained within each subclass and then combined by weighting by the proportion of observations in each subclass.
- OUTCOME REGRESSION, using the PS as a covariate
  - A great attention should be paid to the PS specification.
- Double robust methods combining the PS and the IPTW

Whatever the approach, always check the balance of covariates

- In practice, and it's advisable to check the balance for higher order of variables.(*Austin & Stuart, 2015*)

DIRECT STANDARDISATION

1. Fit the outcome model

```
OutM <- glm(surv30 ~ swang1 + age +factor(sex) +factor(race) +edu +factor(income) +factor(ninsclas) +factor(cat1
```

2. Averaging the exposure effect over the covariate distribution of the standard population

```
Risk_A1 <- predict(OutM, newdata=A1, type="response")
Risk_A0 <- predict(OutM, newdata=A0, type="response")
RD.30ds <- mean(Risk_A1) - mean(Risk_A0)
RR.30ds <- mean(Risk_A1) / mean(Risk_A0)

round(cbind(RD30, RD.30ds, RR.30, RR.30ds), 3)
```

```
##         RD30 RD.30ds RR.30 RR.30ds
## [1,] -0.053  -0.055 0.923    0.92
```

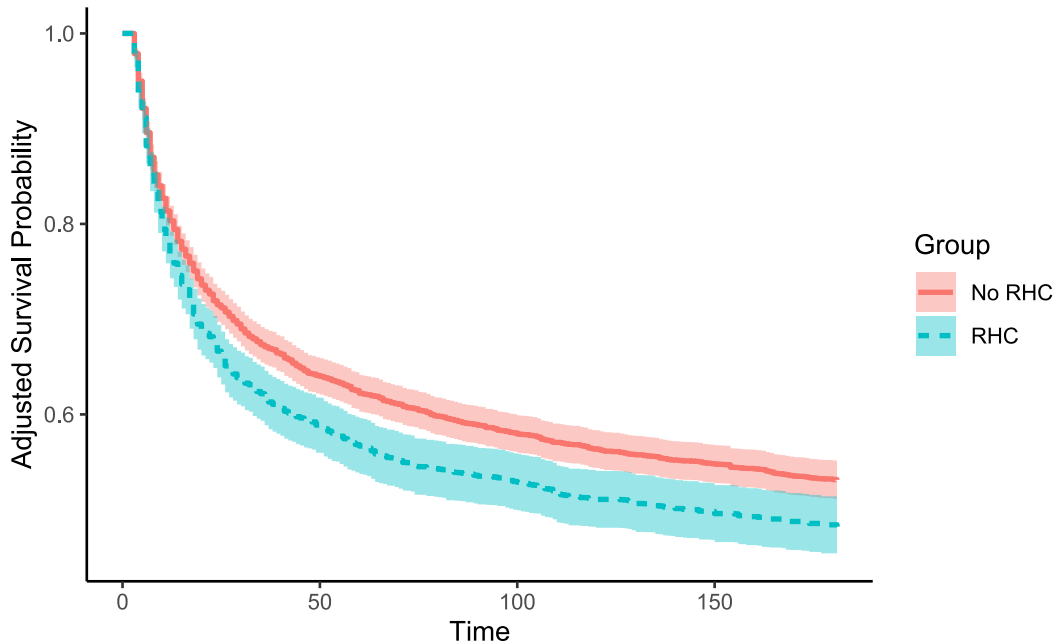
Bootstrapping for confidence intervals.

When the exposure is discrete, models saturated and positivity holds, IPTW = direct standardization. (*Page 24 of Hernan and Robins Causal Inference What If*)

NOTES

We could have performed a survival analysis instead (the idea was to mimic the *JAMA* paper).

```
library(adjustedCurves)
rhc$group <- as.factor(rhc$swang1); rhc$event <- ifelse(rhc$death=="Yes",1,0)
surv <- adjustedsurv(data=rhc, variable="group", ev_time="surv", event="event",
                         method="iptw_km", treatment_model=rhc$SIPTW, conf_int=TRUE)

plot(surv, conf_int=TRUE, linetype=TRUE, max_t=181)
```



CLOSING REMARKS

Assumptions to compute causal effects with observational data:

Consistency: treatment levels correspond to well-defined interventions.

Conditional Exchangeability: the treatment (PS)/outcome (DS) models include the correct set of covariates.

Positivity: there are treated and untreated individuals in all covariate strata.

No interference: an individual's PO doesn't depend on another individual's treatment.

Correct model specification: the functional form of the covariates and chosen link function are correct.

No measurement errors: treatment, outcome and covariates are correctly measured.

CLOSING REMARKS (2)

- Unless the covariates contain very few categorical variables, misspecification of the exposure (PS) or outcome (Direct standirsation) models is likely
    - Misspecification of models could introduce sizable bias
    - With the availability of machine learning tools, we can relax the modelling assumptions.
- The SUPPORT study was a multi-center study
    - The variable indicating the center was not available in the dataset
    - Look at *Langworthy, B., et al 2023* for an overview of propensity score matching methods for clustered data.
- All the methods were covered can be extended for continuous exposures.

References

Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015 Dec 10;34(28):3661-79

Cinelli C, Forney A and Pearl J. A crash course in good and bad controls. Sociological Methods & Research. Published online: May 20, 2022

Cole, S. R., and Hernán, M. A. (2008), "Constructing Inverse Probability Weights for Marginal Structural Models," American Journal of Epidemiology, 168, 656-664.

Cunningham, S. (2021). Causal inference. In Causal Inference. Yale University Press. Chpt 4.

Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC

Lash T. L. VanderWeele T. J. Haneause S. & Rothman K. J. (2021). Modern epidemiology (Fourth). Lippincott Williams & Wilkins. Chpt 3 & 14.

Langworthy, B., Wu, Y., & Wang, M. (2023). An overview of propensity score matching methods for clustered data. Statistical Methods in Medical Research, 32(4), 641-655.