

CAUSAL ESTIMATION - TP

DAVID HAJAGE david.hajage@aphp.fr

DÉPARTEMENT DE SANTÉ PUBLIQUE

JUIN 2024

- 1 Presentation of the data
- 2 Naive analysis of the mortality
- 3 Comparison of treated and untreated
- 4 Standardization
- 5 Propensity score weighting

1

PRESENTATION OF THE DATA

- Patients with liver disease hospitalized for encephalopathy
- Treatment of interest: fecal transplantation
- Cohort study that prospectively included patients
- Primary objective: to evaluate the effectiveness of fecal transplantation for the treatment of hepatic encephalopathy
- Primary endpoint: mortality

```
install.packages(c("gtsummary", "Hmisc", "ggplot2",  
                  "survey", "cobalt", "marginaleffects"))
```

```
load("df.RData")
head(df)
```

```
##          id  age sex  sex.f tobacco  tobacco.f alcohol alcohol.f  bili hiv
## 752527  1 68.50  0  Male      0 Non-smoker      2 Excessive 27.235  0
## 488554  2 72.35  1 Female      0 Non-smoker      2 Excessive 288.305  0
## 67985   3 48.84  1 Female      0 Non-smoker      2 Excessive 48.427  0
## 250417  4 62.61  1 Female      1 Smoker        2 Excessive 24.315  0
## 827567  5 65.46  1 Female      1 Smoker        1 Moderate 98.830  0
## 273278  6 57.39  1 Female      0 Non-smoker      1 Moderate 2.626  0
##          hiv.f trt      trt.f status status.f
## 752527    No  0 Unexposed      1 Deceased
## 488554    No  0 Unexposed      1 Deceased
## 67985     No  0 Unexposed      1 Deceased
## 250417    No  0 Unexposed      0 Alive
## 827567    No  0 Unexposed      1 Deceased
## 273278    No  0 Unexposed      0 Alive
```

Variable	Label	Unit/Coding
id	Identification Number	
age	Age	Year
sex	Sex	0: Male, 1: Female
tobacco	Tobacco	0: Smoker, 1: Non-smoker
alcohol	Alcohol	0: Never, 1: Moderate, 2: Excessive
bili	Bilirubin	$\mu\text{mol/l}^{-1}$
hiv	HIV	0: No, 1: Yes
status	Vital Status	0: Alive, 1: Deceased

Categorical variables have already been recoded as factor (.f at the end of the variable name).

```
summary(df[, c("age", "sex.f", "tobacco.f", "alcohol.f",
               "bili", "hiv.f", "trt.f", "status.f")])
```

```
##      age          sex.f          tobacco.f          alcohol.f
## Min.   :35.18   Male   : 929   Non-smoker: 803   Never      :617
## 1st Qu.:55.48   Female:1071   Smoker    :1197   Moderate   :962
## Median :60.12                                     Excessive:421
## Mean   :60.05
## 3rd Qu.:64.80
## Max.   :85.71
##      bili          hiv.f          trt.f          status.f
## Min.   :  0.214   No :1909   Unexposed:1537   Alive      :1113
## 1st Qu.: 11.079   Yes:  91   Exposed  : 463   Deceased:  887
## Median : 30.498
## Mean   :  95.164
## 3rd Qu.: 77.719
## Max.   :4542.734
```



```
# install.packages("gtsummary")
library(gtsummary)
tbl_summary(df[, c("age", "sex.f", "tobacco.f", "alcohol.f",
                  "bili", "hiv.f", "trt.f", "status.f")],
            statistic = all_continuous() ~ "{mean} ({sd})"
          )
```

Characteristic	N = 2,000
age	60 (7)
sex.f	
Male	929 (46%)
Female	1,071 (54%)
tobacco.f	
Non-smoker	803 (40%)
Smoker	1,197 (60%)
alcohol.f	
Never	617 (31%)
Moderate	962 (48%)
Excessive	421 (21%)
bili	95 (263)
hiv.f	91 (4.6%)
trt.f	
Unexposed	1,537 (77%)
Exposed	463 (23%)
status.f	
Alive	1,113 (56%)
Deceased	887 (44%)

2

NAIVE ANALYSIS OF THE MORTALITY

- Compare mortality between the 2 treatment groups (using description followed by a simple logistic regression)

Characteristic	Unexposed, N = 1,537	Exposed, N = 463
status.f		
Alive	845 (55%)	268 (58%)
Deceased	692 (45%)	195 (42%)

```
summary(glm(status ~ trt.f, data = df, family = "binomial"))
```

```
##  
## Call:  
## glm(formula = status ~ trt.f, family = "binomial", data = df)  
##  
## Coefficients:  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.19975    0.05127  -3.896 9.77e-05 ***  
## trt.fExposed -0.11824    0.10718  -1.103    0.27  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##    Null deviance: 2747.0  on 1999  degrees of freedom  
## Residual deviance: 2745.8  on 1998  degrees of freedom  
## AIC: 2749.8  
##  
## Number of Fisher Scoring iterations: 3
```

- What do you think?

- What do you think?
- Explore the effect of other variables on mortality

Characteristic	Alive, N = 1,113	Deceased, N = 887
age	58 (7)	62 (7)

```
summary(glm(status ~ age, data = df, family = "binomial"))
```

```
##  
## Call:  
## glm(formula = status ~ age, family = "binomial", data = df)  
##  
## Coefficients:  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -5.000528  0.434642  -11.51  <2e-16 ***  
## age         0.079212  0.007148   11.08  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##      Null deviance: 2747.0  on 1999  degrees of freedom  
## Residual deviance: 2610.5  on 1998  degrees of freedom  
## AIC: 2614.5  
##  
## Number of Fisher Scoring iterations: 4
```

The interpretation of this log odds ratio is: for each additional year, the log odds ratio increases by 0.079.

Characteristic	Alive, N = 1,113	Deceased, N = 887
bili	44 (58)	160 (380)

```
summary(glm(status ~ bili, data = df, family = "binomial"))
```

```
##  
## Call:  
## glm(formula = status ~ bili, family = "binomial", data = df)  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.6770725  0.0588254 -11.51  <2e-16 ***  
## bili         0.0064506  0.0006239  10.34  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##    Null deviance: 2747.0  on 1999  degrees of freedom  
## Residual deviance: 2539.9  on 1998  degrees of freedom  
## AIC: 2543.9  
##  
## Number of Fisher Scoring iterations: 6
```

The interpretation of this log odds ratio is: for each additional unit of bilirubin, the risk of death increases by 0.006.

Characteristic	Alive, N = 1,113	Deceased, N = 887
tobacco.f		
Non-smoker	494 (44%)	309 (35%)
Smoker	619 (56%)	578 (65%)

```
summary(glm(status ~ tobacco.f, data = df, family = "binomial"))
```

```
##  
## Call:  
## glm(formula = status ~ tobacco.f, family = "binomial", data = df)  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)   -0.46919    0.07253  -6.469 9.86e-11 ***  
## tobacco.fSmoker  0.40066    0.09277   4.319 1.57e-05 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##    Null deviance: 2747.0  on 1999  degrees of freedom  
## Residual deviance: 2728.2  on 1998  degrees of freedom  
## AIC: 2732.2  
##  
## Number of Fisher Scoring iterations: 4
```

Characteristic	Alive, N = 1,113	Deceased, N = 887
alcohol.f		
Never	382 (34%)	235 (26%)
Moderate	526 (47%)	436 (49%)
Excessive	205 (18%)	216 (24%)

```
summary(glm(status ~ alcohol.f, data = df, family = "binomial"))
```

```
##  
## Call:  
## glm(formula = status ~ alcohol.f, family = "binomial", data = df)  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)    -0.4858     0.0829  -5.860 4.62e-09 ***  
## alcohol.fModerate  0.2982     0.1052   2.834 0.00459 **  
## alcohol.fExcessive 0.5381     0.1280   4.204 2.62e-05 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##    Null deviance: 2747.0  on 1999  degrees of freedom  
## Residual deviance: 2728.5  on 1997  degrees of freedom  
## AIC: 2734.5  
##  
## Number of Fisher Scoring iterations: 4
```

Characteristic	Alive, N = 1,113	Deceased, N = 887
hiv.f	33 (3.0%)	58 (6.5%)

```
summary(glm(status ~ hiv.f, data = df, family = "binomial"))
```

```
##  
## Call:  
## glm(formula = status ~ hiv.f, family = "binomial", data = df)  
##  
## Coefficients:  
##             Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -0.26450    0.04618  -5.728 1.02e-08 ***  
## hiv.fYes     0.82843    0.22288   3.717 0.000202 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##    Null deviance: 2747.0  on 1999  degrees of freedom  
## Residual deviance: 2732.5  on 1998  degrees of freedom  
## AIC: 2736.5  
##  
## Number of Fisher Scoring iterations: 4
```

- Most variables appear to be prognostic

3

COMPARISON OF TREATED AND UNTREATED

```
tbl_summary(df[, c("age", "sex.f", "tobacco.f", "alcohol.f",  
                  "bili", "hiv.f", "trt.f")],  
            statistic = all_continuous() ~ "{mean} ({sd})",  
            by = "trt.f"  
) %>% add_p()
```

Characteristic	Unexposed, N = 1,537	Exposed, N = 463	p-value
age	60 (7)	60 (7)	0.6
sex.f			0.4
Male	722 (47%)	207 (45%)	
Female	815 (53%)	256 (55%)	
tobacco.f			0.010
Non-smoker	641 (42%)	162 (35%)	
Smoker	896 (58%)	301 (65%)	
alcohol.f			<0.001
Never	441 (29%)	176 (38%)	
Moderate	746 (49%)	216 (47%)	
Excessive	350 (23%)	71 (15%)	
bili	77 (208)	155 (389)	<0.001
hiv.f	91 (5.9%)	0 (0%)	<0.001

- What do you think?

- What do you think?
- Several variables are significantly unbalanced between treated and untreated subjects

- What do you think?
- Several variables are significantly unbalanced between treated and untreated subjects
- The `hiv` variable is extreme in this regard: there are no HIV-positive subjects treated with fecal transplantation
- What would you do?

- What do you think?
- Several variables are significantly unbalanced between treated and untreated subjects
- The hiv variable is extreme in this regard: there are no HIV-positive subjects treated with fecal transplantation
- What would you do?
- Being HIV-positive seems to be a contraindication for transplantation
→ positivity problem → exclusion of HIV-positive patients

```
df <- df[df$hiv == 0, ]
```

- Example for age (quantitative variable)

■ Example for age (quantitative variable)

```
m0 <- with(df[df$trt == 0, ], mean(age)) # moy chez les non exposés
m1 <- with(df[df$trt == 1, ], mean(age)) # moy chez les exposés
v0 <- with(df[df$trt == 0, ], var(age)) # var chez les non exposés
v1 <- with(df[df$trt == 1, ], var(age)) # var chez les exposés
(m1 - m0)/sqrt((v1 + v0)/2)
```

```
## [1] -0.04164837
```

- Example for sex (categorical variable)

■ Example for sex (categorical variable)

```
m0 <- with(df[df$trt == 0, ], mean(sex)) # moy chez les non exposés
m1 <- with(df[df$trt == 1, ], mean(sex)) # moy chez les exposés
v0 <- with(df[df$trt == 0, ], var(sex)) # var chez les non exposés
v1 <- with(df[df$trt == 1, ], var(sex)) # var chez les exposés
(m1 - m0)/sqrt((v1 + v0)/2)
```

```
## [1] 0.04926628
```

- Load the `cobalt` library for an easier calculation :

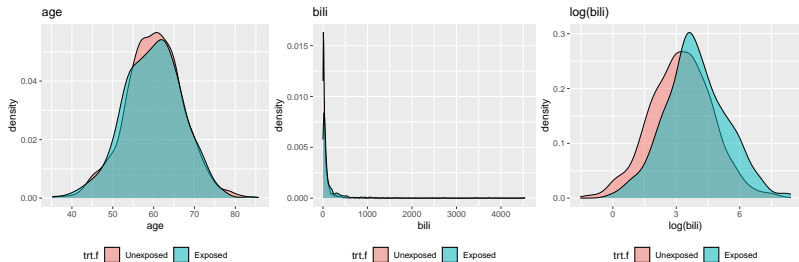
```
library(cobalt)
diffs <- bal.tab(trt ~ age + sex.f + tobacco.f + alcohol.f + bili, data = df,
                 binary = "std")
diffs
```

```
## Balance Measures
##
##          Type Diff.Un
## age          Contin. -0.0416
## sex.f_Female   Binary  0.0493
## tobacco.f_Smoker Binary  0.1341
## alcohol.f_Never Binary  0.1939
## alcohol.f_Moderate Binary -0.0380
## alcohol.f_Excessive Binary -0.1848
## bili          Contin.  0.2441
##
## Sample sizes
##      Control Treated
## All    1446    463
```

- Some variables have severe imbalance (`bili` for example), others have moderate imbalance (`alcohol.f`)

- Some variables have severe imbalance (`bili` for example), others have moderate imbalance (`alcohol.f`)
- In fact, the imbalance between treated and untreated groups can be inspected across the entire distribution

```
# install.packages("ggplot2")
library(ggplot2)
ggplot(data = df, aes(x = age, fill = trt.f)) + geom_density(alpha = 0.5) +
  ggtitle("age") + theme(legend.position="bottom")
ggplot(data = df, aes(x = bili, fill = trt.f)) + geom_density(alpha = 0.5) +
  ggtitle("bili") + theme(legend.position="bottom")
ggplot(data = df, aes(x = log(bili), fill = trt.f)) + geom_density(alpha = 0.5) +
  ggtitle("log(bili)") + theme(legend.position="bottom")
```



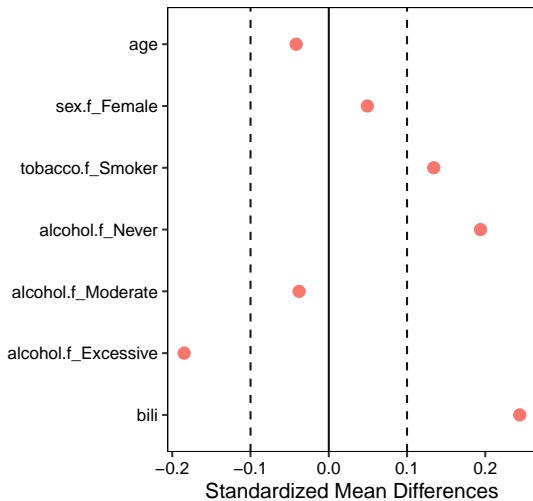
(We observe that bilirubin has a log-normal distribution)

```
df$logbili <- log(df$bili)
```

- It's common practice to display standardized differences on a plot

```
love.plot(diffs, thresholds = c(m = 0.1))
```

Covariate Balance



- Some variables are imbalanced (bili, and alcohol in particular)
- All variables appear to be prognostic of mortality
- Which variables will you consider?

4

STANDARDIZATION

```
## model among trt = 1
mod1 <- glm(status ~ age + sex.f + tobacco.f + alcohol.f + logbili,
            data = subset(df, trt == 1), family = binomial)
## model among trt = 0
mod0 <- glm(status ~ age + sex.f + tobacco.f + alcohol.f + logbili,
            data = subset(df, trt == 0), family = binomial)
```

```
## status predicted if trt = 1
y1ate <- predict(mod1, newdata = df, type = "response")
## status predicted if trt = 0
y0ate <- predict(mod0, newdata = df, type = "response")
```

```
mean(y1ate) - mean(y0ate) ## Risk difference ATE
```

```
## [1] -0.08804777
```

```
log(mean(y1ate) / mean(y0ate)) ## log(RR) ATE
```

```
## [1] -0.2191475
```

```
log((mean(y1ate)/(1-mean(y1ate))) / (mean(y0ate)/(1-mean(y0ate)))) ## log(OR) ATE
```

```
## [1] -0.3669941
```



```
## status predicted if trt = 1
y1att <- predict(mod1, newdata = subset(df, trt == 1), type = "response")
## status predicted if trt = 0
y0att <- predict(mod0, newdata = subset(df, trt == 1), type = "response")
```

```
mean(y1att) - mean(y0att) ## Risk difference ATT
```

```
## [1] -0.05420141
```

```
log(mean(y1att) / mean(y0att)) ## log(RR) ATT
```

```
## [1] -0.1210609
```

```
log((mean(y1att)/(1-mean(y1att))) / (mean(y0att)/(1-mean(y0att)))) ## log(OR) ATT
```

```
## [1] -0.2193785
```

```
library(margins)
mod <- glm(status ~ trt * (age + sex.f + tobacco.f + alcohol.f + logbili),
           data = df, family = binomial) ## model with interactions
```

```
## Risk difference ATE
```

```
avg_comparisons(mod, variables = "trt")
```

```
##
```

```
## Term          Contrast Estimate Std. Error    z Pr(>|z|)    S  2.5 % 97.5 %
## trt mean(1) - mean(0)  -0.088    0.0235 -3.74  <0.001 12.4 -0.134 -0.0419
```

```
##
```

```
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low
```

```
## Type: response
```

```
## log(RR) ATE
```

```
avg_comparisons(mod, variables = "trt", comparison = "lnratioavg")
```

```
##
```

```
## Term Contrast Estimate Std. Error z Pr(>|z|) S 2.5 %
```

```
## trt ln(mean(1) / mean(0)) -0.219 0.0623 -3.52 <0.001 11.2 -0.341
```

```
## 97.5 %
```

```
## -0.097
```

```
##
```

```
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low
```

```
## Type: response
```

```
## log(OR) ATE
```

```
avg_comparisons(mod, variables = "trt", comparison = "lnoravg")
```

```
##
```

```
## Term Contrast Estimate Std. Error z Pr(>|z|) S 2.5 %
```

```
## trt ln(odds(1) / odds(0)) -0.367 0.1 -3.65 <0.001 11.9 -0.564
```

```
## 97.5 %
```

```
## -0.17
```

```
##
```

```
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low
```

```
## Type: response
```

```
## Risk difference ATT
```

```
avg_comparisons(mod, variables = "trt",  
                newdata = subset(df, trt == 1))
```

```
##
```

```
## Term          Contrast Estimate Std. Error    z Pr(>|z|)  S 2.5 % 97.5 %  
## trt mean(1) - mean(0) -0.0542    0.0245 -2.21  0.0271 5.2 -0.102 -0.00615
```

```
##
```

```
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.low
```

```
## Type: response
```

```
## log(RR) ATT
```

```
avg_comparisons(mod, variables = "trt", comparison = "lnratioavg",  
                newdata = subset(df, trt == 1))
```

```
##
```

```
## Term          Contrast Estimate Std. Error    z Pr(>|z|)  S 2.5 %  
## trt ln(mean(1) / mean(0)) -0.121    0.0562 -2.15  0.0312 5.0 -0.231  
## 97.5 %  
## -0.0109
```

```
##
```

```
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.lower, conf.upper
```

```
## Type: response
```

```
## log(OR) ATT
```

```
avg_comparisons(mod, variables = "trt", comparison = "lnoravg",  
                newdata = subset(df, trt == 1))
```

```
##
```

```
## Term Contrast Estimate Std. Error z Pr(>|z|) S 2.5 %
```

```
## trt ln(odds(1) / odds(0)) -0.219 0.0999 -2.2 0.028 5.2 -0.415
```

```
## 97.5 %
```

```
## -0.0237
```

```
##
```

```
## Columns: term, contrast, estimate, std.error, statistic, p.value, s.value, conf.lo
```

```
## Type: response
```

5

PROPENSITY SCORE WEIGHTING

- The exposure is binary, so we can estimate the propensity score using logistic regression
- Since all variables appear to be prognostic, we include all variables in the model

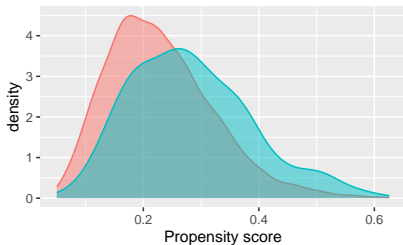
```
mod.trt <- glm(trt ~ age + sex.f + tobacco.f + alcohol.f + logbili,
              data = df, family = "binomial")
summary(mod.trt)
```

```
##
## Call:
## glm(formula = trt ~ age + sex.f + tobacco.f + alcohol.f + logbili,
##      family = "binomial", data = df)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.820024    0.505249  -3.602 0.000315 ***
## age           -0.006039    0.007876  -0.767 0.443261
## sex.fFemale    0.094588    0.110417   0.857 0.391642
## tobacco.fSmoker 0.248305    0.113930   2.179 0.029297 *
## alcohol.fModerate -0.314444    0.121818  -2.581 0.009844 **
## alcohol.fExcessive -0.685826    0.162349  -4.224 2.4e-05 ***
## logbili        0.312401    0.037919   8.239 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2115.1  on 1908  degrees of freedom
## Residual deviance: 2016.6  on 1902  degrees of freedom
## ATC: 2030.6
```

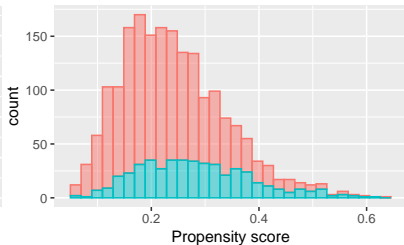
```
df$score <- predict(mod.trt, type = "response")
```

COMPARISON OF THE PROPENSITY SCORE DISTRIBUTION BETWEEN TREATED AND UNTREATED

```
ggplot(data = df, aes(x = score, color = trt.f, fill = trt.f)) +
  geom_density(alpha = 0.5) +
  xlab("Propensity score") +
  theme(legend.position="bottom")
ggplot(data = df, aes(x = score, color = trt.f, fill = trt.f)) +
  geom_histogram(alpha = 0.5) +
  xlab("Propensity score") +
  theme(legend.position="bottom")
```



trt.f ■ Unexposed ■ Exposed



trt.f ■ Unexposed ■ Exposed

COMPARISON OF THE PROPENSITY SCORE DISTRIBUTION BETWEEN TREATED AND UNTREATED

- What do you think?

COMPARISON OF THE PROPENSITY SCORE DISTRIBUTION BETWEEN TREATED AND UNTREATED

- What do you think?
- The distributions of propensity scores for treated and untreated subjects are not perfectly overlapping
- However, their overlap appears adequate, which is reassuring regarding the positivity assumption

- We will create two types of weights: an “ATE” weight and an “ATT” weight
- “ATE” weight

```
df$wate <- ifelse(df$trt == 1, 1/df$score, 1/(1 - df$score))
```

- We will create two types of weights: an “ATE” weight and an “ATT” weight
- “ATE” weight

```
df$wate <- ifelse(df$trt == 1, 1/df$score, 1/(1 - df$score))
```

- “ATT” weight : no weighting for the treated, untreated weighted by $score/(1 - score)$

- We will create two types of weights: an “ATE” weight and an “ATT” weight
- “ATE” weight

```
df$wate <- ifelse(df$trt == 1, 1/df$score, 1/(1 - df$score))
```

- “ATT” weight : no weighting for the treated, untreated weighted by $score/(1 - score)$

```
df$watt <- ifelse(df$trt == 1, 1, df$score/(1 - df$score))
```

- We will create two types of weights: an “ATE” weight and an “ATT” weight
- “ATE” weight

```
df$wate <- ifelse(df$trt == 1, 1/df$score, 1/(1 - df$score))
```

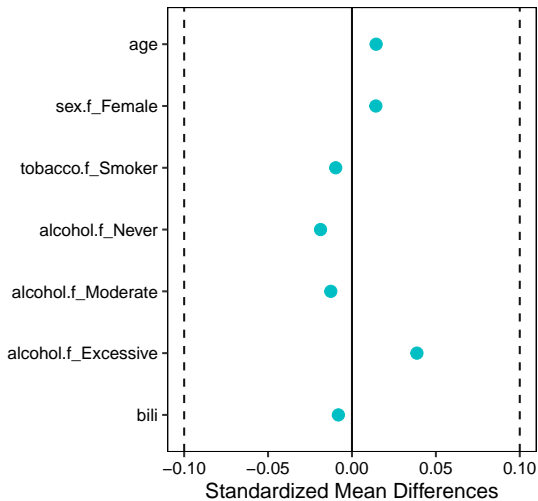
- “ATT” weight : no weighting for the treated, untreated weighted by $score/(1 - score)$

```
df$watt <- ifelse(df$trt == 1, 1, df$score/(1 - df$score))
```

- Examine the balance after applying the different weights (you should use the `std.diff` function with the option `weights = "wate"` or `"watt"`)

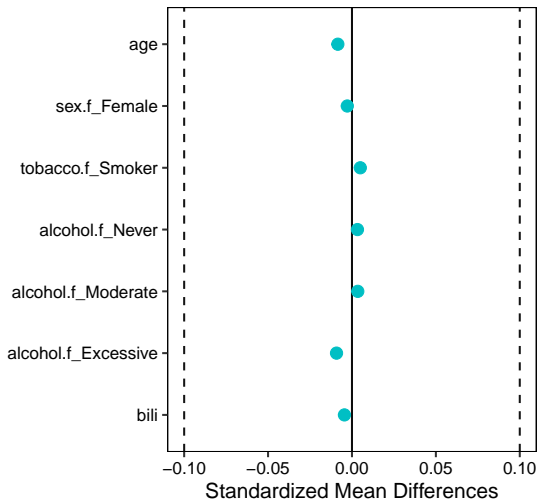
```
diffs <- bal.tab(trt ~ age + sex.f + tobacco.f + alcohol.f + bili,  
               data = df,  
               weights = "wate", binary = "std")  
love.plot(diffs, thresholds = c(m = 0.1))
```

Covariate Balance



```
diffs <- bal.tab(trt ~ age + sex.f + tobacco.f + alcohol.f + bili,  
                data = df,  
                weights = "watt", binary = "std")  
love.plot(diffs, thresholds = c(m = 0.1))
```

Covariate Balance



■ ATE

```
m1 <- with(subset(df, trt == 1), weighted.mean(status, wate))
m0 <- with(subset(df, trt == 0), weighted.mean(status, wate))
m1 - m0 ## Risk difference ATE
```

```
## [1] -0.08294736
```

```
log(m1/m0) ## log RR ATE
```

```
## [1] -0.203498
```

```
log(((m1)/(1-m1))/((m0)/(1-m0))) ## log RR ATE
```

```
## [1] -0.3440832
```

■ ATT

```
m1 <- with(subset(df, trt == 1), weighted.mean(status, watt))  
m0 <- with(subset(df, trt == 0), weighted.mean(status, watt))  
m1 - m0 ## Risk difference ATE
```

```
## [1] -0.06696335
```

```
log(m1/m0) ## log RR ATE
```

```
## [1] -0.1475533
```

```
log(((m1)/(1-m1))/((m0)/(1-m0))) ## log RR ATE
```

```
## [1] -0.2704971
```



```
# install.packages("survey")
library(survey)
## Risk difference ATE
summary(svyglm(status ~ trt.f, family = binomial(link = "identity"),
               svydesign(data = df, ids = ~ 1, weight = ~ wate)))

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
##
## Call:
## svyglm(formula = status ~ trt.f, design = svydesign(data = df,
##           ids = ~1, weight = ~wate), family = binomial(link = "identity"))
##
## Survey design:
## svydesign(data = df, ids = ~1, weight = ~wate)
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.45049    0.01326   33.98 < 2e-16 ***
## trt.fExposed -0.08295    0.02738   -3.03  0.00248 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Log RR ATE

```
summary(svyglm(status ~ trt.f, family = binomial(link = "log"),  
             svydesign(data = df, ids = ~ 1, weight = ~ wate)))
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
##
```

```
## Call:
```

```
## svyglm(formula = status ~ trt.f, design = svydesign(data = df,  
##     ids = ~1, weight = ~wate), family = binomial(link = "log"))
```

```
##
```

```
## Survey design:
```

```
## svydesign(data = df, ids = ~1, weight = ~wate)
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -0.79743    0.02942 -27.101 < 2e-16 ***  
## trt.fExposed -0.20350    0.07151  -2.846  0.00448 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1.000524)
```

```
##
```

```
## Number of Fisher Scoring iterations: 6
```

Log OR ATE

```
summary(svyglm(status ~ trt.f, family = binomial(link = "logit"),
               svydesign(data = df, ids = ~ 1, weight = ~ wate)))
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
##
```

```
## Call:
```

```
## svyglm(formula = status ~ trt.f, design = svydesign(data = df,
##   ids = ~1, weight = ~wate), family = binomial(link = "logit"))
```

```
##
```

```
## Survey design:
```

```
## svydesign(data = df, ids = ~1, weight = ~wate)
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.19870    0.05355  -3.711 0.000212 ***
## trt.fExposed -0.34408    0.11613  -2.963 0.003085 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1.000524)
```

```
##
```

```
## Number of Fisher Scoring iterations: 4
```

Risk difference ATT

```
summary(svyglm(status ~ trt.f, family = binomial(link = "identity"),  
             svydesign(data = df, ids = ~ 1, weight = ~ watt)))
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
##
```

```
## Call:
```

```
## svyglm(formula = status ~ trt.f, design = svydesign(data = df,  
##       ids = ~1, weight = ~watt), family = binomial(link = "identity"))
```

```
##
```

```
## Survey design:
```

```
## svydesign(data = df, ids = ~1, weight = ~watt)
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  0.48813    0.01523  32.060  <2e-16 ***  
## trt.fExposed -0.06696    0.02754  -2.431   0.0151 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1.000524)
```

```
##
```

```
## Number of Fisher Scoring iterations: 3
```

Log RR ATT

```
summary(svyglm(status ~ trt.f, family = binomial(link = "log"),
               svydesign(data = df, ids = ~ 1, weight = ~ watt)))
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
##
```

```
## Call:
```

```
## svyglm(formula = status ~ trt.f, design = svydesign(data = df,
##   ids = ~1, weight = ~watt), family = binomial(link = "log"))
```

```
##
```

```
## Survey design:
```

```
## svydesign(data = df, ids = ~1, weight = ~watt)
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.71717    0.03119  -22.99  <2e-16 ***
## trt.fExposed -0.14755    0.06279   -2.35  0.0189 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1.000524)
```

```
##
```

```
## Number of Fisher Scoring iterations: 6
```

Log OR ATE

```
summary(svyglm(status ~ trt.f, family = binomial(link = "logit"),
               svydesign(data = df, ids = ~ 1, weight = ~ watt)))
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
##
```

```
## Call:
```

```
## svyglm(formula = status ~ trt.f, design = svydesign(data = df,
##   ids = ~1, weight = ~watt), family = binomial(link = "logit"))
```

```
##
```

```
## Survey design:
```

```
## svydesign(data = df, ids = ~1, weight = ~watt)
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -0.04749    0.06094  -0.779    0.436
```

```
## trt.fExposed -0.27050    0.11215  -2.412    0.016 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1.000524)
```

```
##
```

```
## Number of Fisher Scoring iterations: 4
```

THANKS FOR WATCHING!

