UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Confounding and propensity score

Catrin Wessman, catrin.Wessman@gu.se

Lund

October 13, 2016.

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Content

- Causality and confounding
- Neyman-Rubin counterfactual framework
- A first introduction of PS

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

One philosopher's view on causality

David Hume (1711-1776):

"We may define a cause to be an object followed by another, and where all the objects similar to the first, are followed by object similar to the second. Or, in other words, where, if the first object had not been, the second never had existed."

Ref: http://www.ucl.ac.uk/~uctytho/dfwCauseHume.htm

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

A more probabilistic view of causality

- A causal relationship between two variable must have a temporal order
- The two variables should be empirically correlated
- **The correlation cannot be explained by a third variable, which is correlated with both variables (a confounder)**.
- Hence, statistical association is NOT enough to conclude causality.

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Neyman-Rubin counterfactual framework of causality

- Observational data could be treated as missing data; If a patient get treatment A, then the patient has one potential outcome – and the counterfactual outcome is missing.

Individual (i)	Treatment A (Y_{1i})	Treatment B (Y_{0i})
1	4.5	-
2	5.5	-
3	-	1.2
4	3.4	-
5	-	1.3
6	-	1.4
7	4.3	-

$$Y_i = T_i \cdot Y_{1i} + (1 - T_i)Y_{0i}$$

$$T_i = \begin{cases} 1 \; if \; assigned \; to \; treatment \; A \\ 0 \; if \; assigned \; to \; treatment \; B \end{cases}$$

$$Y_{ji} = outcome \; if \; taking \; treatment \; j \; for \; indiviudal \; i$$

The standard estimator for the average effect

- We cannot estimate this at an individual level, since we have missing observations – but, we can estimate this at an average level (since we can impute the missing observations from similar subjects that got the opposite treatment).

- The treatment effect, τ, is given by;

$$\bar{\tau} = E[Y_1|T = 1] - E[Y_0|T = 0]$$

where *W* is 1 or 0, if patient got treatment or not and *Y* is the response.

Some issues with this framework

- We need to be able to assume Stable Unit Treatment Value Assumption (SUTVA) and
- The ignorable treatment assignment assumption.

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

SUTVA – Stable Unit Treatment Value Assumption

"The potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there is no different forms or versions of each treatment level, which lead to different potential outcomes."

Ref: "Causal inference for statistics, social, and biomedical sciences – an introduction", Imbens, Rubin, Cambridge University Press, 2015

SUTVA – An Example of violation of SUTVA

Assume we want to test the effect of a new drug treatment against depression. Two patients (i1 and i2) are involved in the study and they are best friends.

	i1=c| i2=t	i1=t| i2=t	i1=c| i2=c	i1=t| i2=c
i1's score	9	5	21	14

The lower score, the more healthy is i1.

We get different causal effect of i1:s treatment, depending on i2:s treatment.

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

The ignorable treatment assignment assumption

The prerequisite is that given all variables, the assignment to treatment is independent of the outcome. Hence, all individual have the same probability (given the variables) to get both treatments.

If not, we are comparing apple and oranges!

This holds for a experiment proper randomized – any difference is due to chance.

Methods to check baseline variables

- Standardized difference

$$\frac{(\bar{x}_{treat} - \bar{x}_{control})}{\sqrt{(s^2_{treat} + s^2_{control})/2}}$$

- Do not use test-based methods

Pros and cons of RCT?

Pros

- SUTVA-assumption might be controlled and adjusted for
- No problem with ignorability
- Golden standard, high internal validity

Cons

- Not always ethical
- Expensive and time consuming
- Low external validity
- Problems with compliance

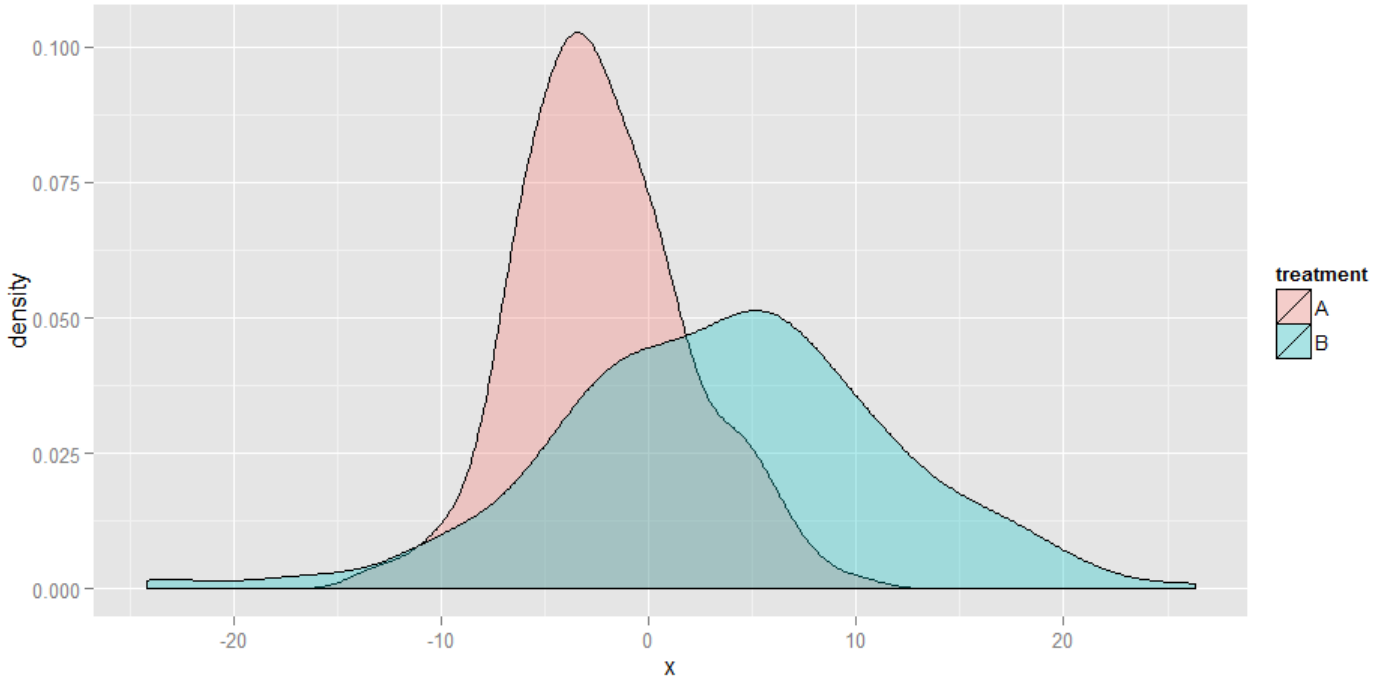
Definition of ATT or ATE

Estimand	Question	Target population
Average Treatment Estimate ATE	What happens if everyone in a given population were exposed to the treatment vs the control?	Whole population
Average Treatment effect of the Treated ATT	What happens to only those persons who actually use the treatment?	Persons actually treated

Note, in a randomized trial, ATT = ATE = ITT

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Examples of ATT or ATE?



UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Linear regression

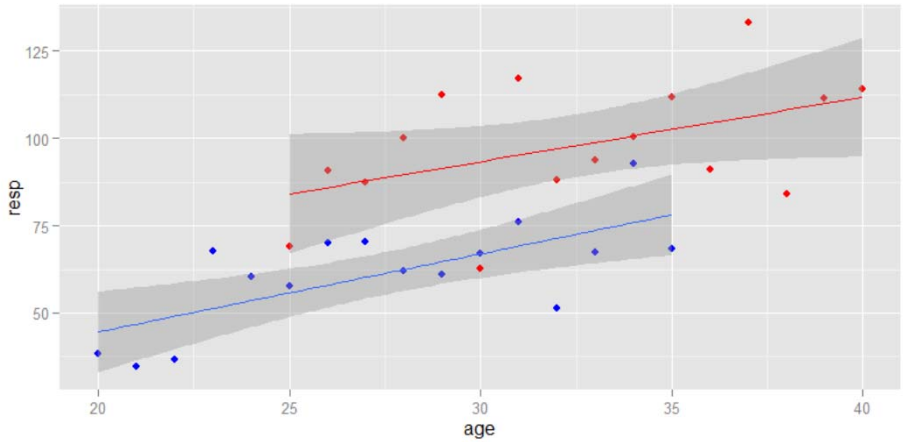
Before analysis

- Check if the dependent variable is continuous
- Plot y against each of the explanatory variables by treatment group. Is it reasonable to expect a linear relation?
- Check if the independent variables do not correlate.
- Check if data is balanced over the range.

After analysis

- Make sure that the residuals are normally distributed.
- Plot the residuals against each of the explanatory variables by treatment group.
- Interpret the coefficients for the dependent variable correctly.

Regression



It is easy to miss serious imbalance between the treatment groups that could decrease the reliability of the regression model unless this is explicitly analysed descriptively. As the variable in the treatment groups become more imbalanced the regression model relies more and more on extrapolation outside the space where data is present (King & Zeng 2005)

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Matching

- Problems of dimensionality
- Categorizing continuous variables
- Problems with generalizability

UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Stratification

- Instead of matching we might stratify data into categories.
- The total effect is estimated by a weighted estimate of the number of elements in each strata.
- The problems of dimensionality is still there!

Propensity score

- Propensity score is ONE way to strengthen the ignorable treatment assumption.
- We estimate the probability that a subject get treatment A, given some variables (X), for each subject (propensity score).
- Then we use this to make sure that our subjects are comparable.

The algorithm

1. Estimate the propensity score

2. Adjust your data, using the propensity score (this can be done by matching, stratification, weighting or with a regression analysis).

3. Assess quality of your adjustment, and redo your propensity score estimate if necessary. This can be done as many times as needed.

4. Estimate your outcome and treatment effect

Estimate propensity score

- We want to estimate the probability get the treatment A or not, in this case it is a dichotomous variable;

$$P(T = T_A | X)$$

- One way to estimate is to estimate a logistic regression, since the treatment is either treatment A or treatment B.

- X is all variables we want and need to consider

- We are allowed to redo this analysis as many times as we like, until we have a regression that give us balance. This is allowed since we are not using the outcome in our estimate

Matching propensity score pairs

- When we have calculate our scores, we pair a subject, who receives treatment A, with a subject, who receives treatment B.
- We might, depending of the variables, X, we put in in our model, get odd ‘couples’.
- We need to have knowledge of all confounders, since the treatment assignment is independent of the potential outcomes conditional on the observed baseline variables, if and only if we have included all important variables

Matching - an example

- The quality of life (pst) nine month after surgery, for 646 patients with mastectomy (1) or lumpectomy (0) are registered in a German Breast Cancer register. Higher scores of pst, reflecting better QoL.

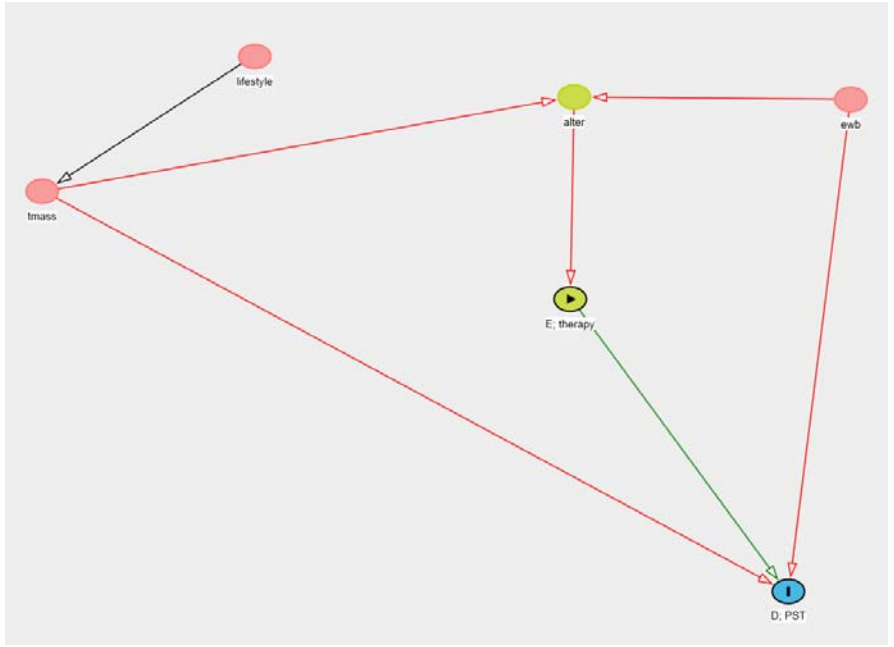
- Other variables are the size of tumor mass (tmass), age (alter) and emotional well-being (ewb).

- We want to estimate the effect in QoL of having a mastectomy instead of having lumpectomy in a patient that is qualified of mastectomy (ATT), adjusted for variables.

Descriptive analysis of data

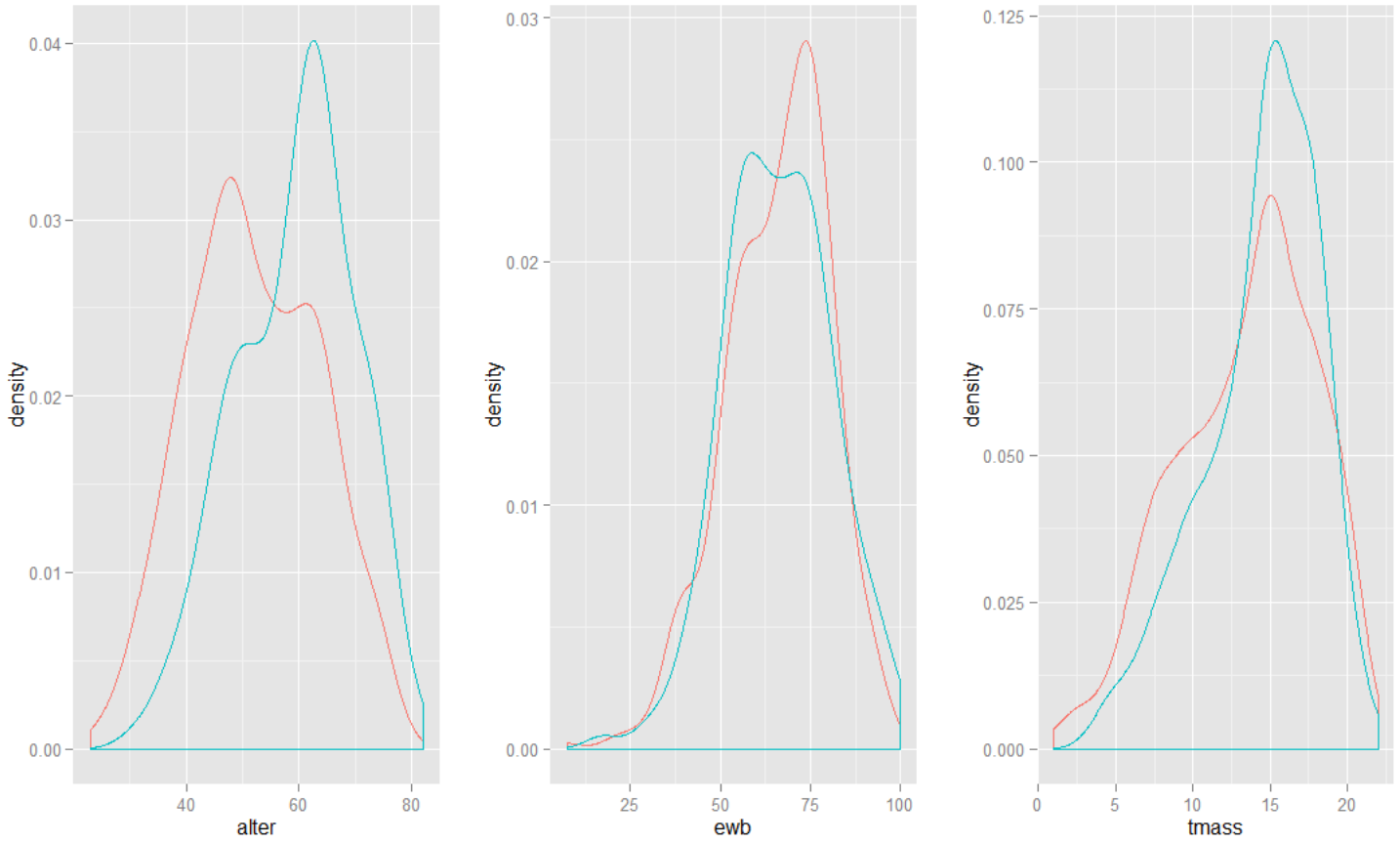
```
      pst   therapy alter  ewb tmass
pst      1.00   -0.05 -0.16  0.74  0.05
therapy -0.05    1.00  0.28  0.00  0.10
alter   -0.16    0.28  1.00 -0.06  0.09
ewb      0.74    0.00 -0.06  1.00  0.02
tmass    0.05    0.10  0.09  0.02  1.00
```

```
> summary(stu1$alter)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  23.00   45.25   54.00   53.91   63.00   82.00
> summary(stu1$ewb)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  7.692  55.770  67.310  66.100  76.920 100.000
> summary(stu1$tmass)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00   11.00   15.00   13.76   17.00   22.00
```



UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

An example of visualization of raw data



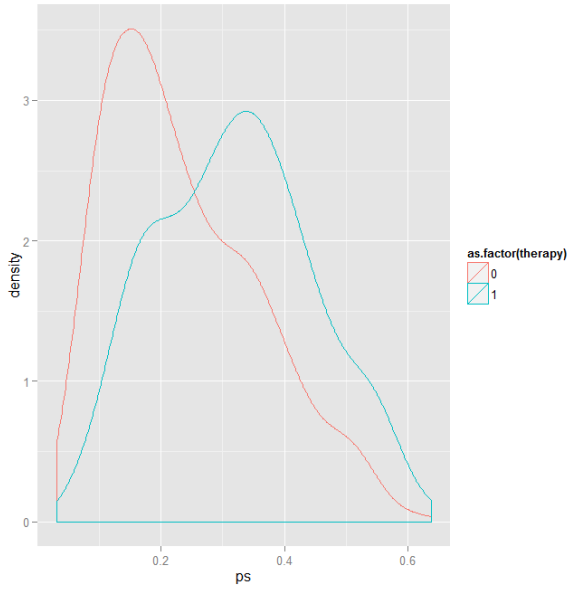
Unmatched data

```
> summary(lm (data=stu1, pst ~ therapy +alter+ewb+tmass))

Call:
lm(formula = pst ~ therapy + alter + ewb + tmass, data = stu1)

Residuals:
    Min      1Q  Median      3Q     Max
-44.657  -5.114   0.493   6.088  28.494

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.47051    2.71666  14.161  < 2e-16 ***
therapy     -0.79844    0.87908  -0.908   0.3641
alter       -0.13829    0.03300  -4.190 3.18e-05 ***
ewb          0.70698    0.02556  27.655  < 2e-16 ***
tmass        0.17927    0.08779   2.042   0.0416 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.368 on 641 degrees of freedom
Multiple R-squared:  0.5575,    Adjusted R-squared:  0.5547
F-statistic: 201.9 on 4 and 641 DF,  p-value: < 2.2e-16
```

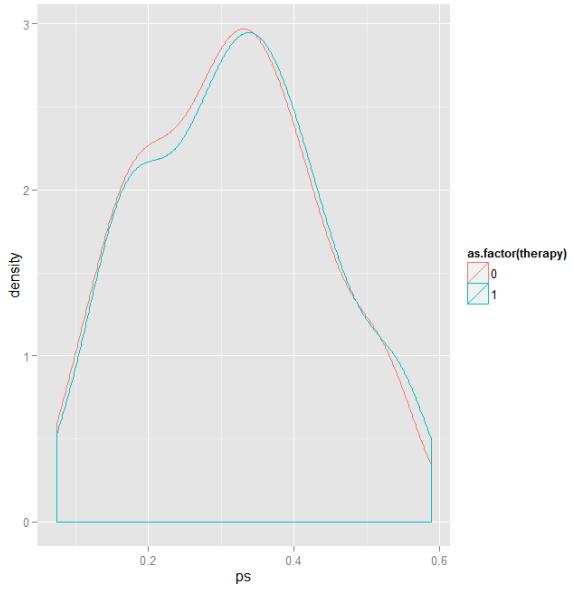
UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

Difference between PS distributions

Before matching



After matching



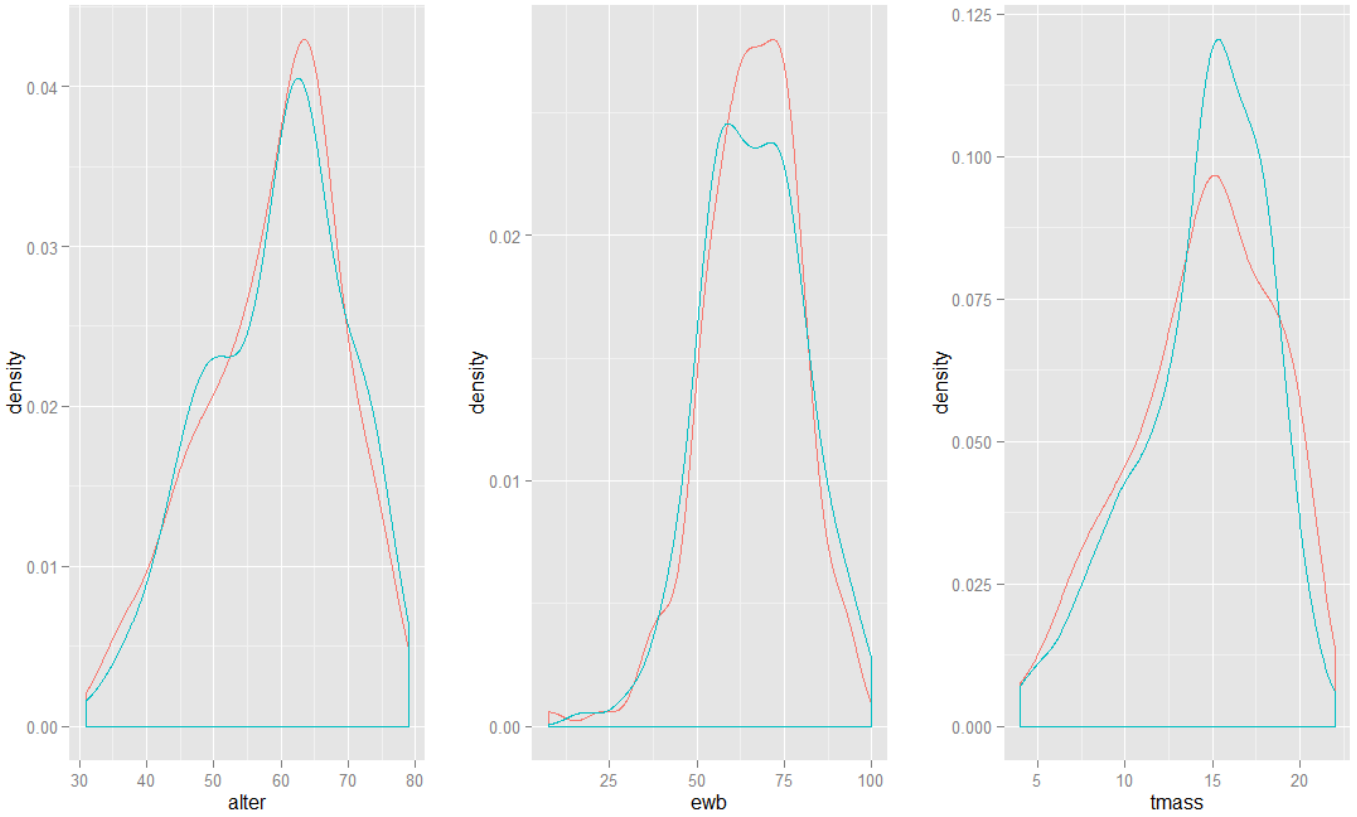
Descriptive after matching

```
> round(cor(tmp1,    method = "pearson"),2)
          pst therapy alter   ewb tmass
pst      1.00   -0.01 -0.20  0.74  0.09
therapy -0.01    1.00  0.02  0.02  0.01
alter   -0.20    0.02  1.00 -0.10 -0.01
ewb      0.74    0.02 -0.10  1.00  0.05
tmass    0.09    0.01 -0.01  0.05  1.00
```

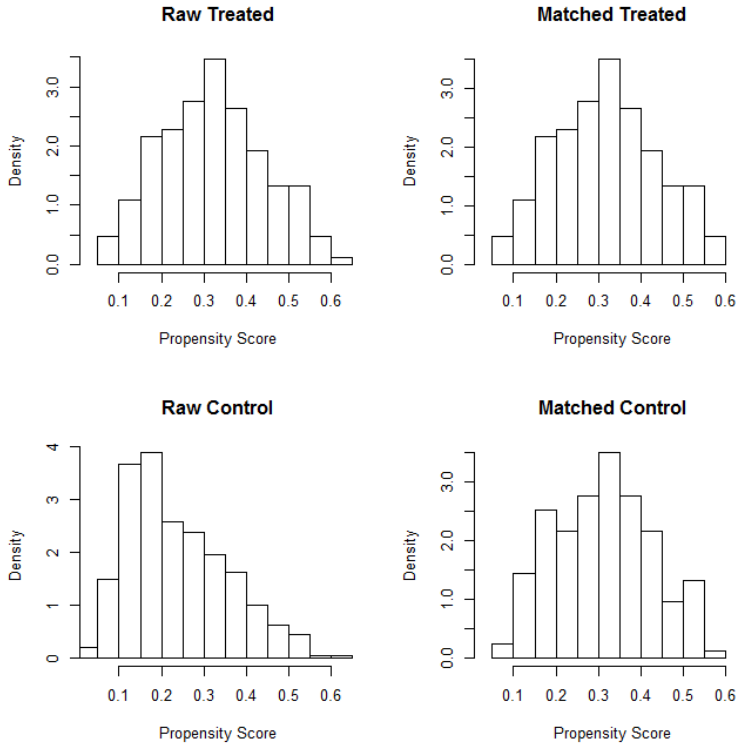
```
> summary(m.data$alter)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  31.00   52.00   61.00   59.05   66.00   79.00
> summary(m.data$ewb)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  7.692  56.130  66.030  66.000  75.000 100.000
> summary(m.data$tmass)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   4.00   12.00   15.00   14.41   17.00   22.00
```

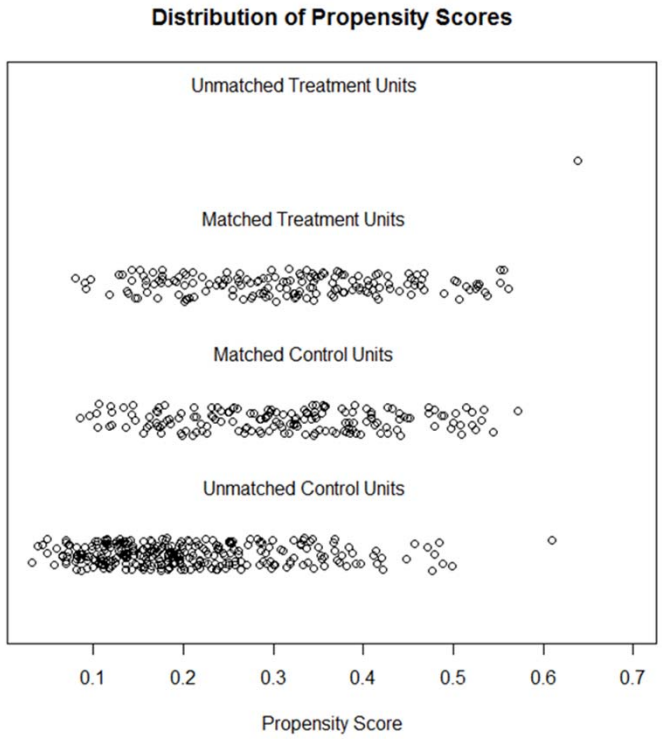
UNIVERSITY OF GOTHENBURG
HEALTH METRICS
Biostatistics, Epidemiology and Health Economics

An example of visualization of matched data

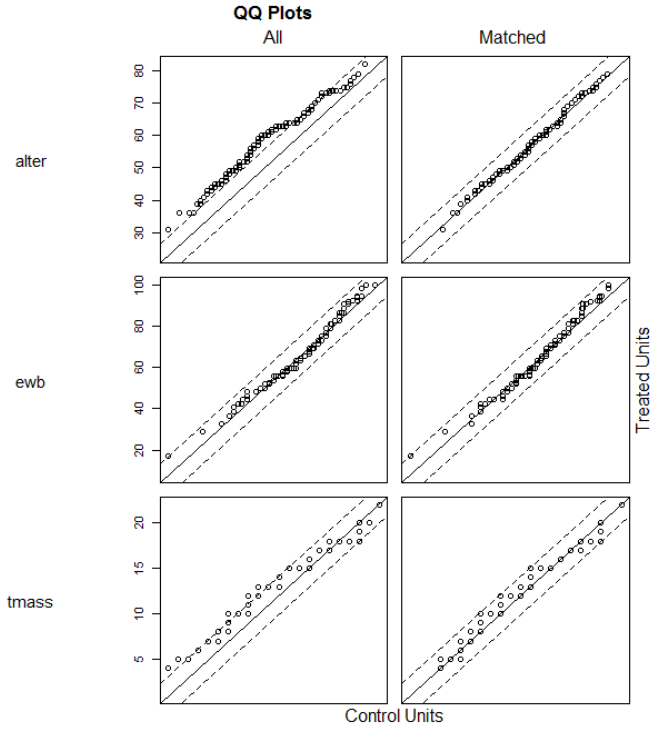


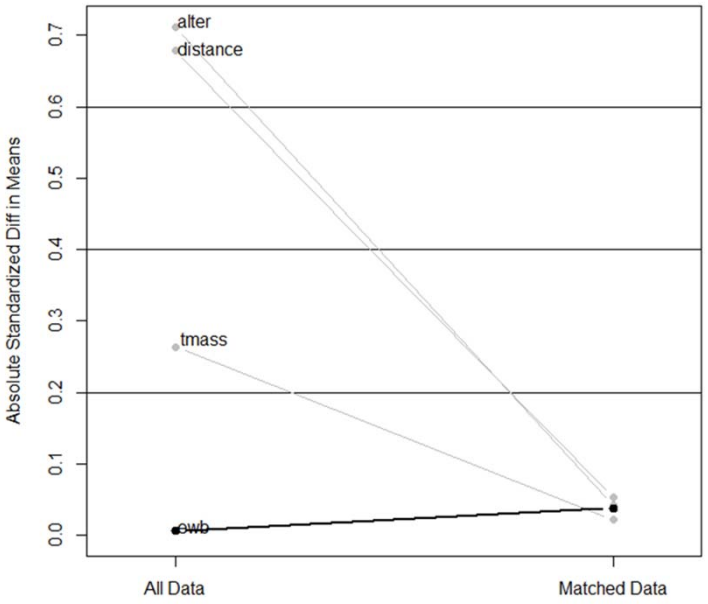
Checking the matched data set





Checking the matched data set, cont…





Result after matching

```
> summary(lm (data=m.data,pst ~ therapy +alter+ewb+tmass))

Call:
lm(formula = pst ~ therapy + alter + ewb + tmass, data = m.data)

Residuals:
     Min       1Q   Median       3Q      Max
-31.1367  -4.9992  -0.4061   5.9788  29.2857

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.54552    4.47717   8.609 3.15e-16 ***
therapy     -0.62436    1.02976  -0.606 0.544725
alter       -0.17158    0.05079  -3.378 0.000817 ***
ewb          0.72691    0.03648  19.924  < 2e-16 ***
tmass        0.19709    0.13428   1.468 0.143150
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.377 on 327 degrees of freedom
Multiple R-squared:  0.5697,    Adjusted R-squared:  0.5645
F-statistic: 108.3 on 4 and 327 DF,  p-value: < 2.2e-16
```

References

Höfler M., “Causal inference based on counterfactuals”, BMC Medical research Methodology, 2005, 5:28

Austin P., “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies”, Multivariate Behav Res. 2011 May; 46(3): 399–424.

Stuart E., “Matching methods for causal inference: A review and a look forward”, Stat Sci. 2010 Feb 1; 25(1): 1–21.