A More Technical Explanation to Propensity Score Matching

The aim of this appendix is to describe the principle of propensity score matching in more technical terms and to establish the link to the linear regression used explanatory in Chapter 2 of Daley et al. (2014).
Matching is a covariate-specific treatment-control comparison, weighted together to provide the average treatment effect[1]. In theory no distributional assumptions is needed and given the *Strongly Ignorable Treatment Assignment* the estimate has a causal interpretation.

Framework

We use Rubin's potential-outcome framework[2]: Outcome for an individual with and without ADHD.

Let $D_j \in [0, 1]$ be an indicator of whether individual j has ADHD or not, and let y_j^1 denote the outcome of interest if the individual has ADHD, whereas y_j^0 is the outcome of interest if the individual does not have ADHD. The impact of having ADHD on outcome y for individual j is then defined as $y_j^1 - y_j^0$. The fundamental evaluation problem is that the same individual with both outcomes is not observed. The observed outcome is given by

$$y_j = \begin{cases} y_j^1 & \text{if} \quad D_j = 1 \\ y_j^0 & \text{if} \quad D_j = 0 \end{cases}$$

Daley et al. (2014) focuses on constructing means and establishing the expected causal change in y for individuals with ADHD (which in the literature is called "the average effect of treatment on the treated", δ_{ATT});

$$\delta_{ATT} = E[y^1 - y^0 \mid D = 1] = E[y^1 \mid D = 1] - E[y^0 \mid D = 1]$$

where $E[. \mid D = 1]$ denotes the expected value of a variable given that the individual has ADHD. The challenge is to find the counterfactual, i.e., $E[y^0 \mid D = 1]$, which is unobserved. The relevant question is; "What is the potential outcome of an individual with ADHD if that same individual did not have ADHD"? This fundamental problem is solved under the *Strongly Ignorable Treatment Assignment* assumption by an estimation technique called matching.

Ignorability of treatment

A simple comparison of observed outcome from an individual with ADHD and from an individual without ADHD, $\Delta \equiv E(y^1 \mid D = 1) - E(y^0 \mid D = 0)$, is likely contaminated by the effects of other variables that are correlated with ADHD and the potential outcome - this is similar to the discussion of omitted variable bias but in the treatment literature

[1] In this study: The average treatment effect of the treated
[2] Rubin (1974)

referred to as selection bias;

$$\begin{aligned}\Delta &\equiv E[y^1 \mid D=1] - E[y^0 \mid D=0] \\ &= E[y^1 \mid D=1] - E[y^0 \mid D=1] + E[y^0 \mid D=1] - E[y^0 \mid D=0] \\ &= \delta_{ATT} + \Lambda\end{aligned}$$

The simple sample mean equals *the average treatment effect (of the treated)* plus *the selection bias*, $\Lambda \equiv E[y^0 \mid D=1] - E[y^0 \mid D=0]$.

If the development of ADHD is random the selection bias is not a problem. But genetics and the environment is known to influence the probability of having ADHD. It is likely that individuals with ADHD would have less educational achievement even in the absent of their ADHD condition because their parents are less educated (possible due to an ADHD condition of the parents). To estimate the effect of ADHD is an econometric challenge because the triggers of ADHD are unobserved (and partly unknown) - the solution is to find a good proxy.
The estimation method requires that (y^0, y^1) are independent of ADHD, conditional on available information. This is referred to as *ignorability of treatment* (Rosenbaum and Rubin (1983)) or *selection on observables*.

Let X be a vector of pre-treatment variables then the *ignorability of treatment* is stated like,

$$\{y^0, y^1\} \perp\!\!\!\perp D \mid X \tag{1}$$

It is important that the distribution of the pretreatment variables is not influenced by the ADHD condition[3].

The assumption in (1) is a bit too strong. Mean independence is enough (for most of the results) and given that this study is interested in the treatment effect on the treated the assumption boils down to:

$$E[y^0 \mid X, D] = E[y^0 \mid X, D=0] = E[y^0 \mid X, D=1] = E[y^0 \mid X]$$

Given this assumption, conditional-on-X comparison of average outcome across ADHD and non-ADHD has a causal interpretation[4],

$$\begin{aligned}\delta_{ATT} \equiv E[y^1 - y^0 \mid D=1] &\overset{*}{=} E\left[E[y^1 \mid X, D=1] - E[y^0 \mid X, D=1] \mid D=1\right] \\ &= \int [E[y^1 \mid X, D=1] - E[y^0 \mid X, D=0]] dF(X \mid D=1) \\ &= \int \Delta_x dF(X \mid D=1),\ \Delta_X \equiv E[y^1 \mid X, D=1] - E[y^0 \mid X, D=0]\end{aligned}$$

In words, δ_{ATT} is obtainable as a weighted average of contrasts between ADHD and matched non-ADHD outcome at each value of X. The weights are given by the density of X among individuals with ADHD.
A simple matching estimator will replace δ_X by the sample difference and combine using the empirical distribution of covariates among individuals with ADHD.

[3] This will be explained later

[4] (*) is by *the law of iterated expectations*

Exogeneity and omitted variable bias

The following section explores the link between selection bias and the standard exogeneity assumption. This is also an introduction to the link between OLS and the matching techniques.

First assume that the treatment effect is constant: $\Delta_x = \delta_{ATT} = \delta_{ATE}$[5] $= y_j^1 - y_j^0 \ \forall \ j$.
Observed outcome is given by

$$y_j = y_j^0 + D_j(y_j^1 - y_j^0)$$

which implies by *the ignorability of treatment*,

$$\begin{aligned} E[y_j \mid X_j, D_j] &= E[y_j^0 \mid X_j, D_j] + E[D_j(y_j^1 - y_j^0) \mid X_j, D_j] \\ &= E[y_j^0 \mid X_j] + \delta_{ATT} D_j \end{aligned}$$

When the treatment effect is constant, $E[y_j \mid X_j, D_j]$ is additive in D_j (with the constant treatment effect as the coefficient) and a function of X_j.
Observed outcome is given by[6]

$$\begin{aligned} y_j &= E[y_j \mid X_j, D_j] + \varepsilon_j, \ E[\varepsilon_j \mid X_j, D_j] \overset{*}{=} 0 \\ &= E[y_j^0 \mid X_j] + \delta_{ATT} D_j + \varepsilon_j, \ E[\varepsilon_j \mid X_j, D_j] = 0 \end{aligned}$$

This is the standard exogeneuity assumption and standard regression methods can be used to estimate δ_{ATT}. So far nothing is assumed about the structure of $E[y_j^0 \mid X_j]$ and nonlinear regression methods can be used if y^0 is assumed to be nonlinear in the parameters and otherwise a flexible functional form could be used.

Suppose, now, that the control-outcome is linear in X_j:

$$y_j^0 = \alpha + X_j'\beta + \varepsilon_j^0, \ E[\varepsilon_j^0 \mid X_j] = 0$$

Then the residual in the linear model

$$y_j = \alpha + X_j'\beta + \delta_R D_j + \varepsilon_j$$

is uncorrelated with X_j and D_j and δ_R is the causal effect[7].

A standard regression framework assumes a constant effect but this is not the case for the matching estimator. The last part of this section introduces the regression framework without the assumption of a constant treatment effect.
For the intuition it is useful to notice the difference between δ_{ATE} and δ_{ATT}. First decompose y_0 and y_1 into their mean and a stochastic part with zero mean; $y_0 = \mu_0 + \eta_0$, where $\mu_0 = E[y_0]$ and $y_1 = \mu_1 + \eta_1$, where $\mu_1 = E[y_1]$. Then

$$y_1 - y_0 = (\mu_1 - \mu_0) + (\eta_1 - \eta_0)$$

[5]The average treatment effect
[6](*) is by construction (*the CEF Decomposition Property*)
[7]The constant treatment effect: $\Delta_x = \delta_R = \delta_{ATT} = \delta_{ATE}$

Taking the conditional-on-X expectation gives

$$\begin{aligned} E[y_1 - y_0 \mid D = 1] &= (\mu_1 - \mu_0) + E(\eta_1 - \eta_0 \mid D = 1) \\ \delta_{ATT} &= \delta_{ATE} + E(\eta_1 - \eta_0 \mid D = 1) \end{aligned}$$

The average treatment effect of the treated equals *the average treatment effect* and *the expected individual-specific gains of those who are treated.*
Also note that by taking the conditional-on-X-and-D-expectation[8]

$$\Delta_x \overset{*}{=} \delta_{ATE}(x) \overset{*}{=} \delta_{ATT}(x) = (\mu_1 - \mu_0) + E[\eta_1 - \eta_0 \mid X = x] = \delta_{ATE} + E[\eta_1 - \eta_0 \mid X = x]$$

the average treatment effect at X equals the *average treatment effect* and the *expected individual-specific gain at* x.

Lets return to the conditional mean,

$$\begin{aligned} E[y_j \mid X_j, D_j] &= E[y_j^0 \mid X_j, D_j] + E[D_j(y_j^1 - y_j^0) \mid X_j, D_j] \\ &= E[y_j^0 \mid X_j] + E[y_j^1 - y_j^0 \mid X_j] \times D_j \end{aligned}$$

Now, $E[y_j^1 - y_j^0 \mid X_j]$ is not constant but a function of X and therefore *the average treatment effect (on the treated)* will be affected by the distribution of X and in general $\delta_{ATT} \neq \delta_{ATE}$.

$$\begin{aligned} E[y_j \mid X_j, D_j] &= E[y_j^0 \mid X_j] + E[y_j^1 - y_j^0 \mid X_j] \times D_j \\ &= E[y_j^0 \mid X_j] + \delta_{ATE} D_j + E[\eta_1 - \eta_0 \mid X_j] \times D_j \end{aligned}$$

$E[y_j \mid X_j, D_j]$ is additive in D_j (with δ_{ATE} as the coefficient), a function of X_j, and an interaction term between D_j and a function of X_j that captures the difference between the expected individual-specific effect of those with ADHD and those without ADHD.
Observed outcome using *the conditional decomposition* is given by

$$\begin{aligned} y_j &= E[y_j \mid X_j, D_j] + \varepsilon_j,\ E[\varepsilon_j \mid X_j, D_j] \overset{*}{=} 0 \\ &= E[y_j^0 \mid X_j] + \delta_{ATE} D_j + E[\eta_1 - \eta_0 \mid X] \times D_j,\ \varepsilon_j \perp\!\!\!\perp D_j \mid X_j \end{aligned}$$

This shows how to correct the simple linear framework by an interaction term that captures the differences in individual-specific gains by groups.
With this link in mind the simple framework in Chapter 2 in Daley et al. (2014) is used for explanatory reasons - but it should be stressed that the matching analysis do not suffer from this simplification.

The link to the standard exogeneity assumption also demonstrates that the study should be concerned by the standard problems of endogeneity:

- Omitted variable bias: Unobserved heterogeneity and selection bias
- Measurement error
- Simultaneity or reversed causality

[8] (*) is by *ignorability of treatment*

In the study the biggest concern is omitted variable bias. In a standard regression formulation the omitted variable bias formula tells that by omitting an important variable, say an unobserved variable U, the coefficient will be biased:

$$\delta_{biased} = \delta_R + \psi' U$$

That is by omitting U the coefficient will pick up the causal effect plus the omitted bias.This is analogous to the selection bias derived earlier:

$$\delta_{biased} = \delta_{ATT} + E[y^0 \mid D = 1] - E[y^0 \mid D = 0]$$

By omitting U the estimate will pick up the causal effect and the selection bias.

Given the enormous amount of data available from Statistics Denmark it is possible to control for many characteristics which makes the assumption of *ignorability of treatment* more plausible. But unobservables are still a concern and as a robustness check the study uses siblings as a matched control group to handle unobserved heterogeneity.

Common support

Both treated and non-treated individuals are needed for an appropriate match:

$$0 < Pr(D = 1 \mid X) < 1 \tag{2}$$

Given that the study is interested in δ_{ATT} it is enough that individuals-without-ADHD-matches can be found:

$$Pr(D = 1 \mid X) < 1$$

For every $x \in X$ there should be a positive probability to find a non-ADHD match. If $Pr(D = 1 \mid X) = 1$ the match is not defined and if $Pr(D = 1 \mid X) = 0$ then zero weight is given when calculating *the average treatment effect*.
If the common support assumption is not satisfied for some values of X this limits the estimate to an average treatment effect for the subgroup of individual with ADHD where $Pr(D = 1 \mid X) < 1$.

Again the amount of data available in this study is a huge advantage - the search for a non-ADHD match is from the entire adult Danish population and therefore relative good matches are guaranteed.

Together assumption (1) and (2) are called *Strongly Ignorable Treatment Assignment*.

Identification and causality

First this section investigates the bias, $\tilde{\Delta} - \delta_{ATT}$, of an estimator that adjust for X when *Strongly Ignorable Treatment Assignment* is not fulfilled. The *net treatment effect* at x is

introduced as[9],

$$\Gamma(x) \equiv E(y^1 \mid X = x) - E(y^0 \mid X = x)$$

and its expectation over X; $\tilde{\Gamma}_X \equiv E\{\Gamma(X)\}$.
It is helpful to split the bias into two components:

$$\delta_{bias} \equiv \tilde{\Delta}_X - \delta_{ATT} = (\tilde{\Delta}_X - \tilde{\Gamma}_X) + (\tilde{\Gamma}_X - \delta_{ATT})$$

$(\tilde{\Delta}_X - \tilde{\Gamma}_X)$ represent the degree to which ADHD alters the population regression due to the non-randomness, in the sense that the population regression is not representative for the ADHD and non-ADHD subpopulation. $(\tilde{\Gamma}_X - \delta_{ATT})$ represent the consequences of inappropriate averaging of a conditional expectation to obtain a marginal expectation - as long as X only contains pre-treatment variables this second component will be zero because the variables cannot be affected by the treatment[10].
If treatment assignment is *strongly ignorable* for y_1 and y_0 given the (pretreatment) variables in X, then appropriate adjustment for X is sufficient to directly estimate *the average treatment effect (on the treated)*, i.e. $(\tilde{\Delta}_X - \tilde{\Gamma}_X)$ will be zero.

In theory all that is needed for identification and causality is *strongly ignorable treatment assignment*. But we still have to be careful. X should be pretreatment variables and the propensity score method well specified. The following two section will explore that further.

Pre- vs- post-treatment control (reversed causality)

This section will explain the importance of conditioning on pretreatment variables. This is not directly contained in the *Strongly Ignorable Treatment Assignment* assumption and therefore it presents a bias on its own.
If a quantity is believed to be unaffected by treatment it can be used without concerns - but even if a quantity is affected by treatment there could still be reasons to believe that this is minor compared to the effect treatment has on outcome. If pretreatment measures of an important quantity is unavailable it is possible more valuable to use a post-treatment control than doing nothing. But in general an estimate that adjust for a confounding variable that has been affected by ADHD is biased.

In the following it is explained what kind of bias the affected variable creates.
First introduce (S^1, S^0) as observable values of a post-treatment confounded variable, and the (X, S)-adjusted treatment difference,

$$\Delta(x, s) \equiv E(y^1 \mid D = 1, S_1 = s, X = x) - E(y^0 \mid D = 0, S_0 = s, X = x)$$

and the average (X, S)-adjusted difference, $\tilde{\Delta}_{X,S} \equiv E\{\Delta(X, S)\}$.
To evaluate the bias we reintroduce the *net treatment difference*:

$$\Gamma(x, s) = E(y^1 \mid S_1 = s, X = x) - E(y^0 \mid S_0 = s, X = x)$$

[9]Rosenbaum (1984). This is not a treatment effect - it is a constructed "optimal" effect without any side effect - for example if ADHD influences family situation the net treatment effect do not include this effect even though it essentially is part of the effect of ADHD

[10]See section **pre- vs. post-variables**

and *the average net treatment effect* $\tilde{\Gamma}_{X,S} = E\{\Gamma(X,S)\}$. Now, lets return to the two-component-bias-formula,

$$\delta_{bias} = \tilde{\Delta}_{X,S} - \delta_{ATT} = (\tilde{\Delta}_{X,S} - \tilde{\Gamma}_{X,S}) + (\tilde{\Gamma}_{X,S} - \delta_{ATT})$$

If $S_1 = S_0$ (no effect of treatment on S) then $\tilde{\Gamma}_{X,S} = \delta_{AAT}$, but we still need *Strongly Ignorable Treatment Assignment*[11] to have $\tilde{\Delta}_{X,S} = \tilde{\Gamma}_{X,S}$.
In general, however, $\tilde{\Delta}_{X,S}$ does not equal the average treatment effect, δ_{ATT}, nor the X-adjusted treatment difference, $\tilde{\Delta}_X$. Only if treatment assignment is *strongly ignorable* for (y^1, y^0) given X then the X-adjusted differences equals the average treatment effect - adjustment for X alone is sufficient to remove bias, but the (X,S)-adjusted difference need not equal δ_{AAT}, i.e. adjustment for (X,S) can introduce a bias that could have been avoided by simply applying the X-adjusted differences. Adjustment for (X,S) instead of adjustment for X alone are justified only when they are unnecessary. Therefore this do not give us a rationale for using post-treatment variables (except for possible efficiency gains).

In some cases the post-treatment variable removes a bias. If U is an unobserved confounding (pretreatment) variable that fulfills the *Strongly Ignorable Treatment Assignment*:

$$\begin{aligned} & (y^1, S^1, y^0, S^0) \perp\!\!\!\perp D \mid (X,U) \\ 0 < \quad & Pr(D = 1 \mid X = x, U = u) \quad < 1 \quad \forall\ (x,u) \in (X,U) \end{aligned}$$

and it is possible to observe U then (X,U)-adjustment is enough to estimate the effect of treatment on (y^1, y^0) and (S^1, S^0). But U is unobservable.
The explanation needs one more definition. (S^1, S^0) is defined as a surrogate (Rosenbaum 1984) for U if

$$y^t \perp\!\!\!\perp (S^t, X),\ t = 0, 1$$

For each subpopulation (individuals with ADHD, individuals without ADHD), the outcome and the posttreatment variable are unrelated given the pretreatment variables.
If (S^1, S^0) is a surrogate for U, then $(\tilde{\Delta}_{X,S} - \tilde{\Gamma}_{X,S})$ is zero. More formally if treatment assignment is *strongly ignorable* for (y^1, S^1, y^0, S^0) given both X and U, and if (S^1, S^0) is a surrogate for U, then $(\tilde{\Delta}_{X,S} = \tilde{\Gamma}_{X,S})$. If in addition (S^1, S^0) is unaffected by the treatment then $\tilde{\Delta}_{X,S} = \delta_{ATT}$.

$\tilde{\Delta}_{X,S}$ equal the average treatment effect if three conditions are simultaneous fulfilled:

- Treatment assignment is *strongly ignorable* for (y^1, S^1, y^0, S^0) given X and U
- (S^1, S^0) is a surrogate for U
- (S^1, S^0) is unaffected by the treatment

The problem is that (S^1, S^0) is often affected by treatment and the simplest approach would be to avoid all adjustment for post-treatment variables when treatment assignment is *strongly ignorable* given X. However, when treatment assignment is not *strongly ignorable* given X, this approach can be unsatisfactory, especially when (S^1, S^0) is thought

[11] $\tilde{\Delta} = \delta_{AAT}$ as showed in previous section

to be closely related to an unmeasured pretreatment variable that is relevant to both the condition of ADHD and outcome variables. In such cases, the X-adjusted treatment difference is possible worse as an approximation to δ_{ATT} than the (X, S)-adjusted difference. Adjustment for a post-treatment confounding variable is often used in absent of an unobserved pretreatment variable.

In Daley et al. (2014) many background variables are used to control for initial differences between individuals with ADHD and individuals without ADHD. We believe that to some extent those variables can substitute for the "direct initial measures" that determine the development of ADHD. For example we expect that family situation is a valuable control variable - the groups differ considerable with respect to this variable. This could either reflect effects of ADHD or inherent unobserved pretreatment differences between the groups, or a combination of the two. We believe in the value of using family situation as a control variable and with our cost-objective in mind we prefer a conservative estimate.

Propensity score matching

The aim of Daley et. al. (2014) is to estimate the cost of the causal effect of ADHD. In doing so we use propensity score matching techniques where individuals are matched on a number of background variables. This leads to a dimensionality problem, and therefore we prefer to use a propensity score method (Rosenbaum and Rubin (1983)) to summarize the vector of characteristics, X, into a single-index variable, the propensity score, $p(X) \equiv Pr(D = 1 \mid X)$. The propensity score essentially measures the probability that the individual has ADHD, given his or her initial conditions. Rosenbaum and Rubin (1983) shows that:

$$\text{If } (y^1, y^0) \perp\!\!\!\perp D \mid X \text{ then}$$

$$(y^1, y^0) \perp\!\!\!\perp D \mid p(X)$$

This reduces the multidimensional problem into a univariate problem but requires that the propensity score can be estimated. In this study we estimate the propensity score using probit models. In this way, it should be noted that, we still rely on some functional form assumption even though this is not strictly required in matching in general.
In practice it is difficult to find an individual with exact same propensity score. Hence, we choose the nearest-neighbor algorithm. That is, we compare every individual with ADHD with the individual without ADHD that has the propensity score closest to the propensity score of the individual with ADHD.
Many of our individuals with ADHD has an exact match based on the propensity score - but for some we find a non-exact (but very close) match.
This introduce a functional form assumption that potentially creates some kind of extrapolation bias. As a robustness check we use exact-matching on some of the variables that is expected to have particular influences on the outcome variables.

Matching vs. OLS

This section will highlight the differences between matching and OLS or perhaps, more accurate, highlight the similarities between the two. Both methods are control strategies and mainly differs in their respective weighting scheme. In our view the main advantage of matching is the explanatory focus on causality and common support.

First, lets investigate the differences in weighting schemes:

- Matching uses the distribution of covariates among the treated to weight covariate-specific estimates into an estimate of *the effect of treatment on the treated*:

$$\begin{aligned}
\delta_{ATT} &\equiv E[y^1 - y^0 \mid D = 1] \\
&= E[E[y^1 - y^0 \mid X, D = 1] \mid D = 1] \\
&= E[E[y^1 \mid X, D = 1] - E[y^0 \mid X, D = 1] \mid D = 1] \\
&= E[E[y^1 \mid X, D = 1] - E[y^0 \mid X, D = 0] \mid D = 1] \\
&= \int \Delta_x dF(X \mid D = 1) \\
&= E[\Delta_X \mid D = 1]
\end{aligned}$$

  In the discrete case the matching estimate can be written as:

$$\delta_{ATT} = \sum_x \Delta_x P(X_j = x \mid D_j = 1)$$

  The weights are proportional to the probability of treatment at each value of the covariates.

- OLS produces a conditional-variance-weighted[12] average of these effects

$$\begin{aligned}
\delta_R &\equiv \frac{Cov(Y, \tilde{D})}{V(\tilde{D})} \\
&= \frac{E[(D - E[D \mid X])Y_i]}{E[(D - E[D \mid X])^2]} \\
&= \frac{E[(D - E[D \mid X])E[Y \mid D, X]]}{E[(D - E[D \mid X])^2]} \\
&= \frac{E[(D - E[D \mid X])^2 \Delta_x]}{E[(D - E[D \mid X])^2]} \\
&= \frac{E[E[(D - E[D \mid X])^2 \mid X]\Delta_x]}{E[E[(D - E[D \mid X])^2 \mid X]}
\end{aligned}$$

  which in the discrete case can be written as:

$$\delta_R = \frac{\sum_x \Delta_x [P(D_j = 1 \mid X_j = x)(1 - P(D_j \mid X_j = x))] P(X_j = x)}{\sum_x [P(D_j = 1 \mid X_j = x)(1 - P(D_j \mid X_j = x))] P(X_j = x)}$$

[12] The OLS-estimate is the maximum-likelihood estimator

The treatment-on-the-treated estimate puts the most weight on covariate cells containing those who are most likely to be treated. In contrast, OLS puts the most weight on covariate cells where the conditional variance of treatment status is largest. This is of little importance if Δ_x does not vary across cells.

With saturated controls both the OLS and matching estimands impose common support. In practice, however, both OLS and matching estimators are implemented using modeling assumptions that involve extrapolation across cells. Matching often combine covariate cells with fewer observation and OLS is typical not a saturated model (continous covariates in any finite sample requires functional form restrictions). The fact that both stratification and functional form approximation can be made increasingly accurate as the sample size grows suggests that also these aspects are similar.

Matching estimators are often used in the literature to evaluate treatment effects. However, the above indicates that this is more or less the same as OLS - we continue the matching-tradition because of the pedagogical focus on causality and common support.

Summary

The following table is included to sum up this appendix and provide an overview when reading Chapter 2 in Daley et al. (2014)

Challenges in causality

		Challenge
1	Violation of *Strongly Ignorable Treatment Assignment*	
(i)	Violation of *Ignorablity of treatment*	Selection bias
(ii)	Violation of *Common support*	Restriction of treatment group
2	Using posttreatment variables	Reversed causality
3	Using propensity score	
(i)	Violation of structural assumption	Small sample bias
(ii)	Non-exact matches	Extrapolation