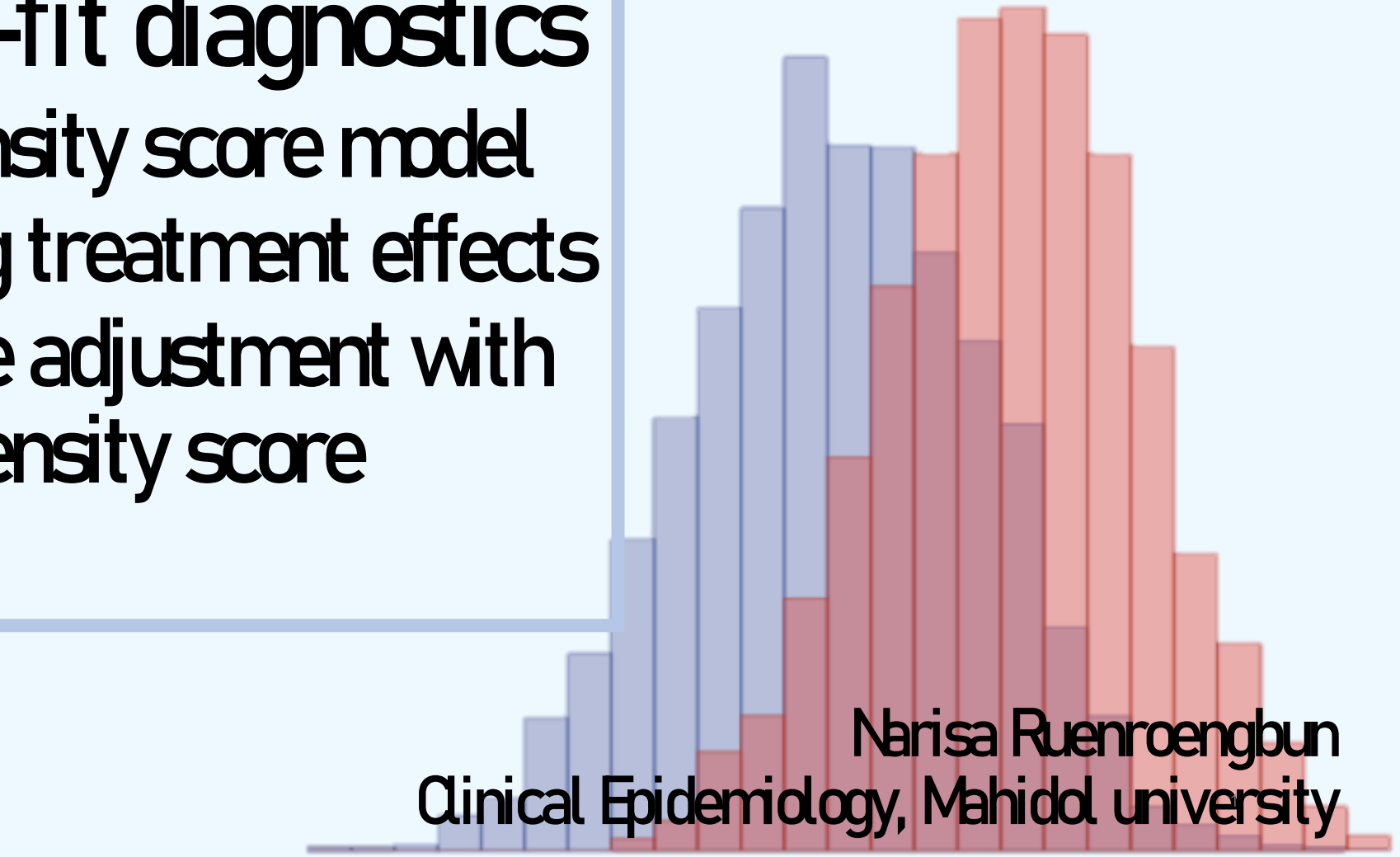
Goodness-of-fit diagnostics

for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score



Narisa Ruenroengbun
Clinical Epidemiology, Mahidol university

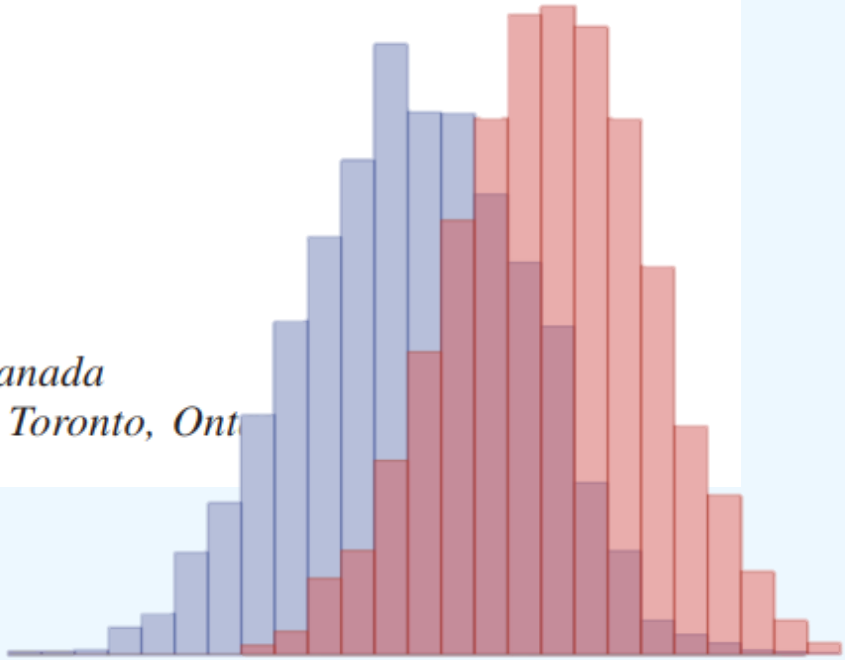
PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2008; **17**: 1202–1217
Published online 29 October 2008 in Wiley InterScience (www.interscience.wiley.com) **DOI**: 10.1002/pds.1673

ORIGINAL REPORT

Goodness-of-fit diagnostics for the propensity score model when estimating treatment effects using covariate adjustment with the propensity score†

Peter C. Austin PhD[1,2,3]*

[1]*Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada*
[2]*Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada*
[3]*Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ont*



Outline

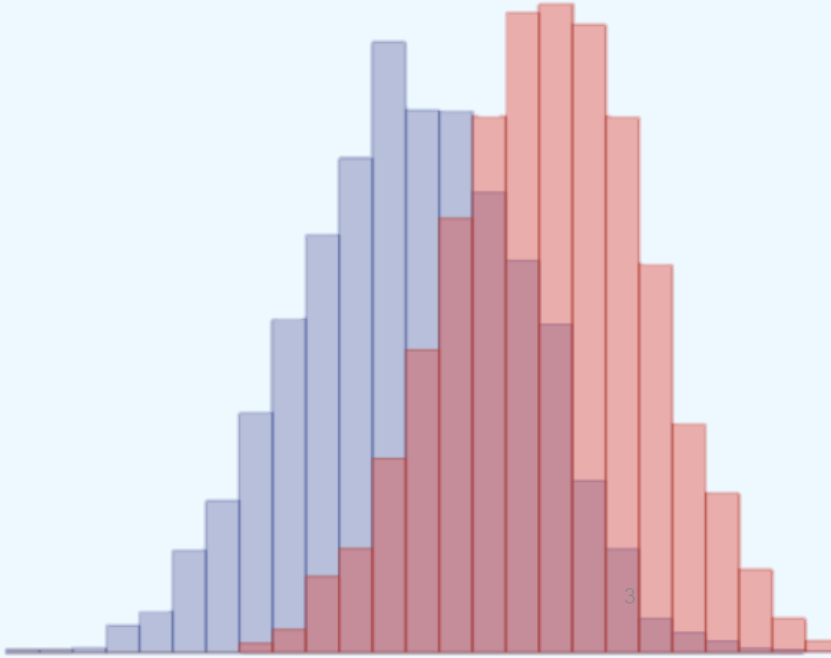
Introduction the difference of propensity score in RCT vs Observational study

Three assumptions used in treatment-effects estimator

Balancing diagnosis for the propensity score model : Covariate adjustment

Quantile regression to compare the distribution

Discussion



Introduction : RCT vs RWD

RCT

No differ systematically between Tx & non Tx subjects in both measured and unmeasured baseline characteristics

Any differences in outcomes can be attributed to the treatment or exposure

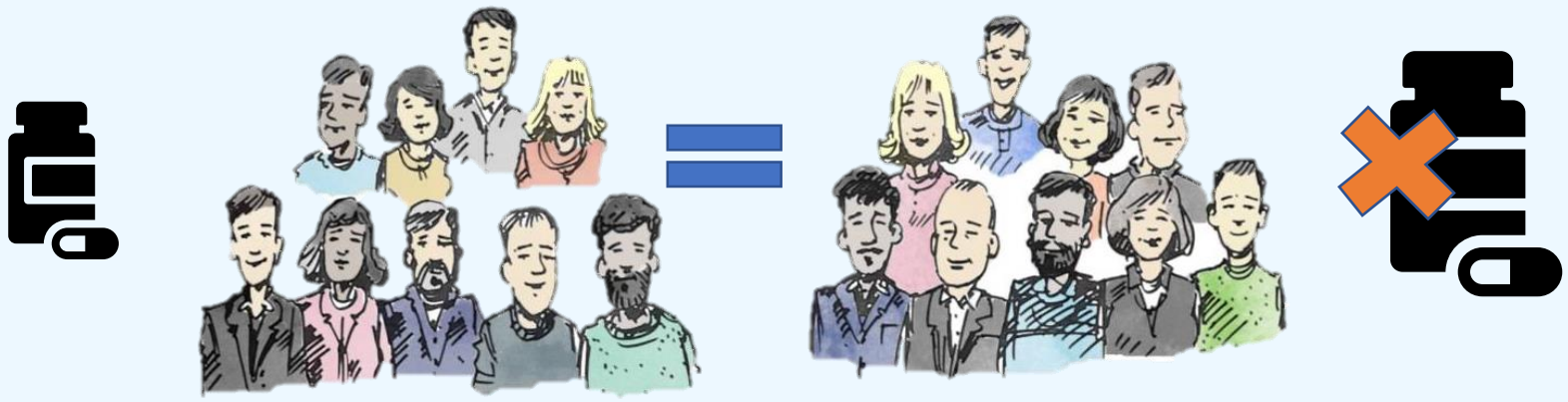
Observational

Treatment selection bias

Unequal the probability of treatment assignment conditional on measured baseline covariates between Tx and non Tx

Propensity score

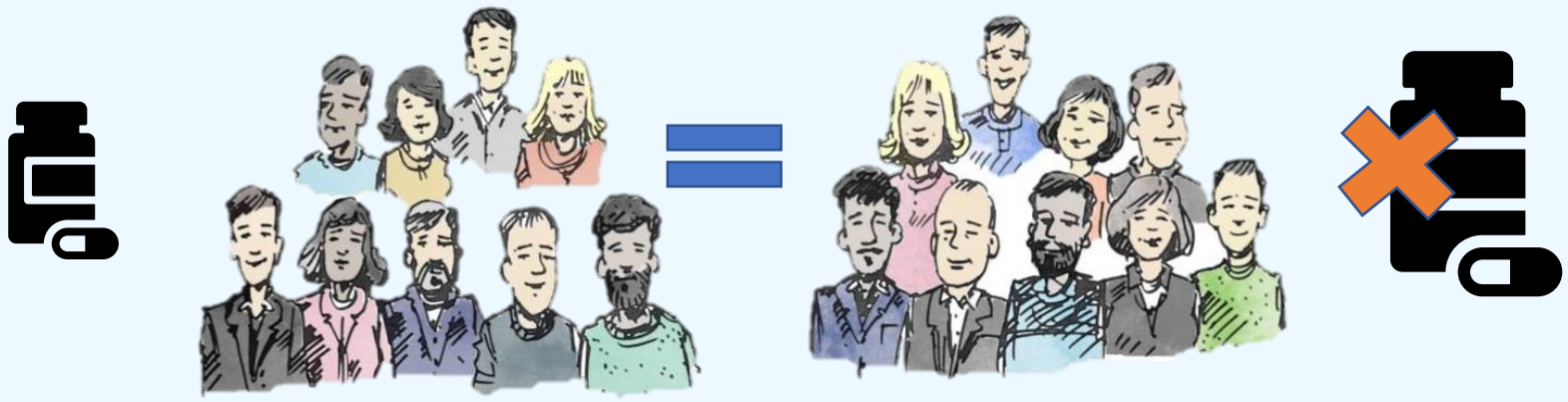
- The probability of treatment assignment conditional on measured baseline covariates.
- Rosenbaum and Rubin purposed a key property
- “ Treatment status is independent of measured baseline covariates”
- “ treated and untreated subjects with the same propensity score will have similar distributions of observed baseline covariates ”



Propensity score

- Three methods of using the propensity score are commonly

1. Covariate adjustment using the propensity score
2. Stratification or subclassification on the propensity score
3. Matching on the propensity score



Propensity score

RCT

The true propensity score is known and is fixed by the design of the experiment

baseline covariates balanced by design

Observational

The true propensity score is unknown and must be estimated using the data

logistic (or probit) regression model in which treatment selection is regressed on measured baseline covariates

Testing the PS have similar distribution of measured baseline covariates

Balancing assumption in TE

Conditional independent (CI)

Treatment model is independent from the potential outcome model.

The CI assumption allows us to estimate the effects by regression-adjustment (RA) methods, inverse probability-weighting (IPW) methods, methods that combine RA and IPW concepts, and matching methods.

Treatment overlap

The overlap assumption ensures that each individual could receive any treatment level

States that individual persons would have positive probability to receive treatments
Similar to RCT in which individual persons would have a chance to receive treatments equally

The independent and identically distributed (i.i.d.)

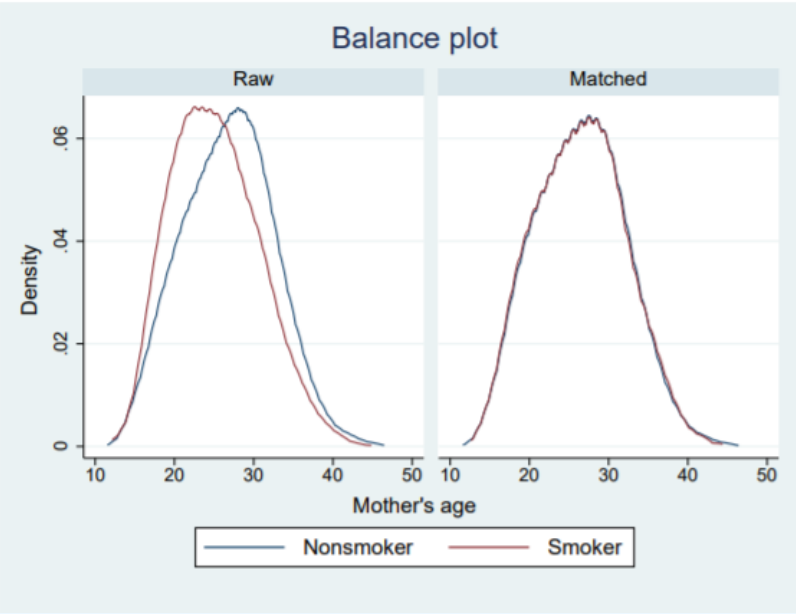
The potential outcomes and the treatment status of each individual are unrelated to the potential outcomes and treatment statuses of all other individuals in the population.

The limitation of testing the assumption in TE

```
. tebalance summarize
Covariate balance summary
                          Raw       Matched
------------------------------------------------
Number of obs =         4,642         9,284
Treated obs   =           864         4,642
Control obs   =         3,778         4,642
------------------------------------------------
```

	Standardized differences		Variance ratio	
	Raw	Matched	Raw	Matched
mmarried	-.5953009	.0014107	1.335944	.9987659
mage	-.300179	-.0120277	.8818025	.9952916
prenatal1	-.3242695	.0333609	1.496155	.9491524
fbaby	-.1663271	-.0117326	.9430944	.9969095

```
tebalance density mage, bwidth(*1.5)
```



It has not been implemented for multivalued treatments

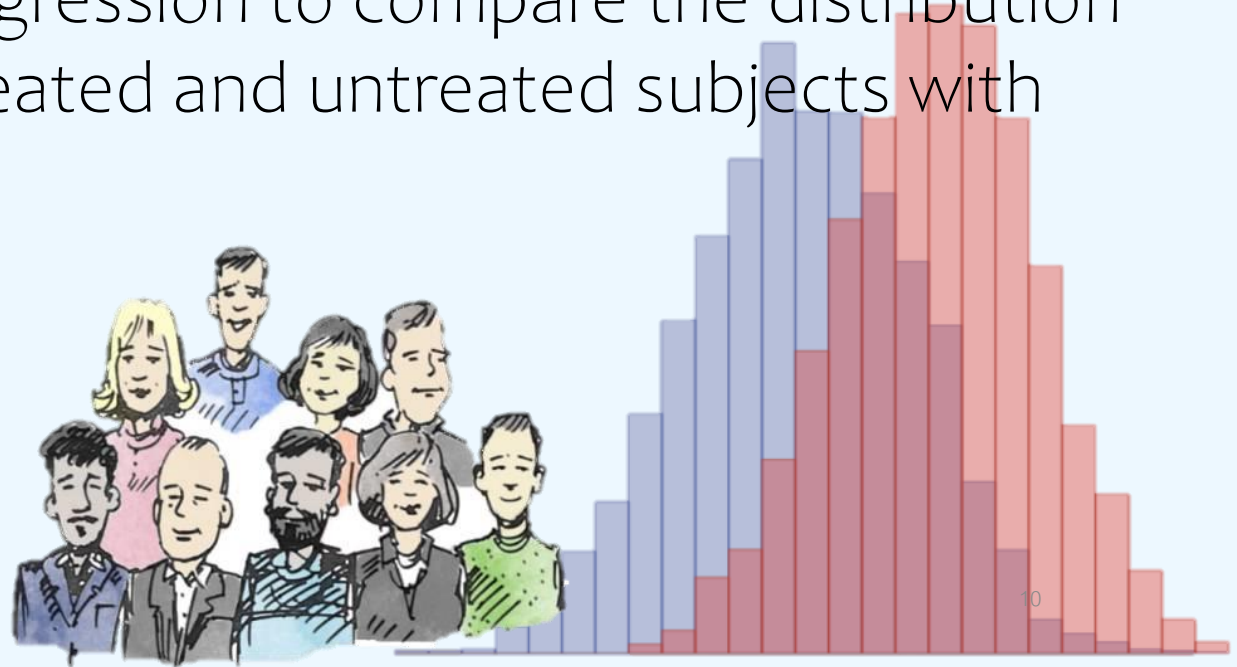
Testing for balancing on propensity score

1

first method extends the standardized difference to the context of covariate adjustment using the propensity score. We refer to this method as the weighted conditional standardized difference.

2

e second method uses quantile regression to compare the distribution of continuous covariates between treated and untreated subjects with similar propensity scores.



Weighted conditional standardized differences

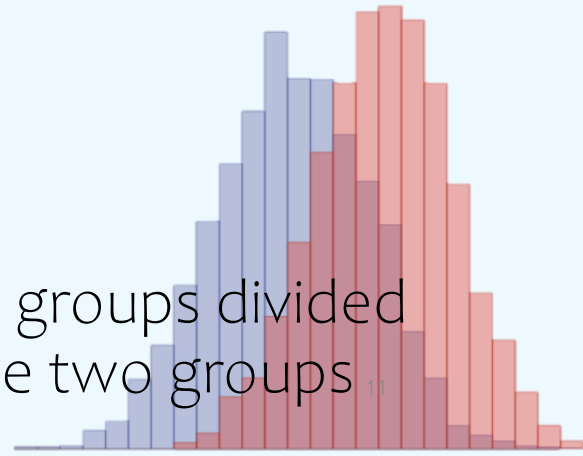
The standardized difference for continuous variables is defined as

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

The standardized difference for dichotomous variables is defined as

$$d = \frac{\hat{p}_{\text{treatment}} - \hat{p}_{\text{control}}}{\sqrt{\frac{\hat{p}_T(1-\hat{p}_T) + \hat{p}_C(1-\hat{p}_C)}{2}}} \qquad (3)$$

The standardized difference is the difference in means between the two groups divided by an estimate of the common standard deviation of that variable in the two groups



BALANCE DIAGNOSTICS FOR THE PROPENSITY SCORE MODEL

The principal idea

Subjects with a similar propensity score, treated and untreated subjects had a similar distribution of baseline covariates.

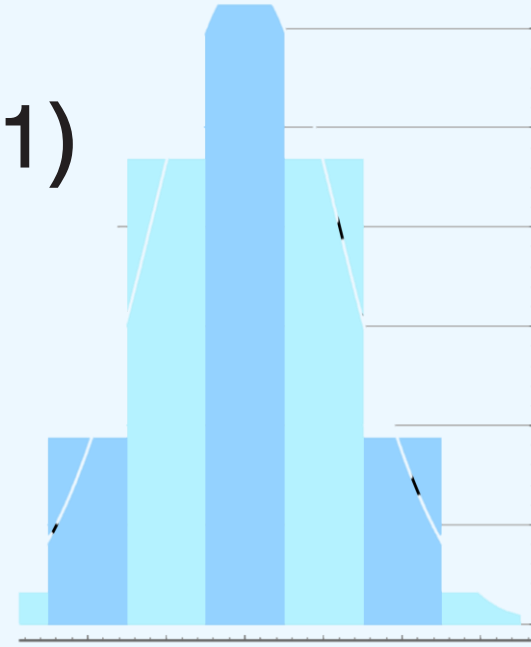
When covariate adjustment using the estimated propensity score is employed, the following regression model is fit to the sample data

$$Y = \alpha_0 + \alpha_1 T + \alpha_2 Z + \varepsilon \qquad (1)$$

Y denotes the outcome

T is an indicator variable denoting treatment selection
(T = 1 denoting treated; T = 0 denoting untreated)

Z denotes the estimated propensity score



BALANCE DIAGNOSTICS FOR THE PROPENSITY SCORE MODEL

The standardized difference for

$$T=1 \quad = \alpha_0 + \alpha_1(1) + \alpha_2 Z + \alpha_3 Z(1) \text{ ----------1}$$

$$T=0 \quad = \alpha_0 + \alpha_1(0) + \alpha_2 Z + \alpha_3 Z(0)$$

$$T=0 \quad = \alpha_0 + \alpha_2 Z \text{ ----------2}$$

$$T1 - To = [\alpha_0 + \alpha_1(1) + \alpha_2 Z + \alpha_3 Z(1)] - \alpha_0 - \alpha_2 Z$$

$$T1 - To = \frac{\alpha_1(1) + \alpha_3 Z}{\sigma}$$

Y denotes the outcome

T is an indicator variable denoting treatment selection (T = 1 denoting treated; T = 0 denoting untreated)

Z denotes the estimated propensity score

BALANCE DIAGNOSTICS

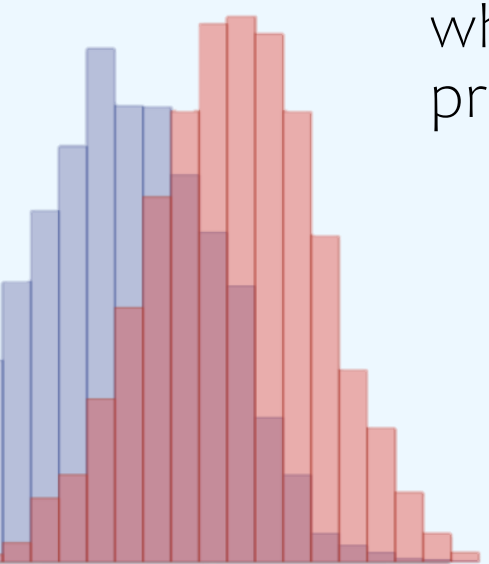
How the conditional standardized absolute difference can be co-computed in practice

$$\int_{z} \frac{|\hat{\alpha}_1 + \hat{\alpha}_3 Z|}{\hat{\sigma}} dZ$$

First, for a given baseline covariate X

$$Y = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 T + \boldsymbol{\alpha}_2 Z + \boldsymbol{\alpha}_3 TxZ$$

one fits the linear regression model described in Equation 1



which X is regressed on an indicator variable denoting treatment status, the propensity score, and the interaction between these two terms.

$$\text{Logit(Pro(X))= 1} = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 T + \boldsymbol{\alpha}_2 Z + \boldsymbol{\alpha}_3 TxZ \quad (4)$$

$$\text{Regress(X)} = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 T + \boldsymbol{\alpha}_2 Z + \boldsymbol{\alpha}_3 TxZ \quad (5)$$

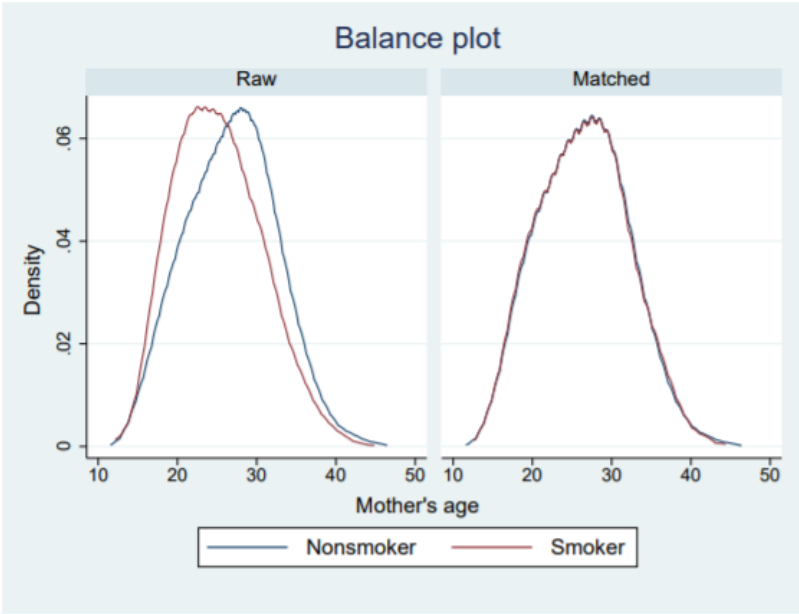
hypothesis: H_0: $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_3 = 0$

The limitation of testing the assumption in TE

```
. tebalance summarize
Covariate balance summary
                          Raw       Matched
------------------------------------------------
Number of obs =         4,642         9,284
Treated obs   =           864         4,642
Control obs   =         3,778         4,642
------------------------------------------------
```

	Standardized differences		Variance ratio	
	Raw	Matched	Raw	Matched
mmarried	-.5953009	.0014107	1.335944	.9987659
mage	-.300179	-.0120277	.8818025	.9952916
prenatal1	-.3242695	.0333609	1.496155	.9491524
fbaby	-.1663271	-.0117326	.9430944	.9969095

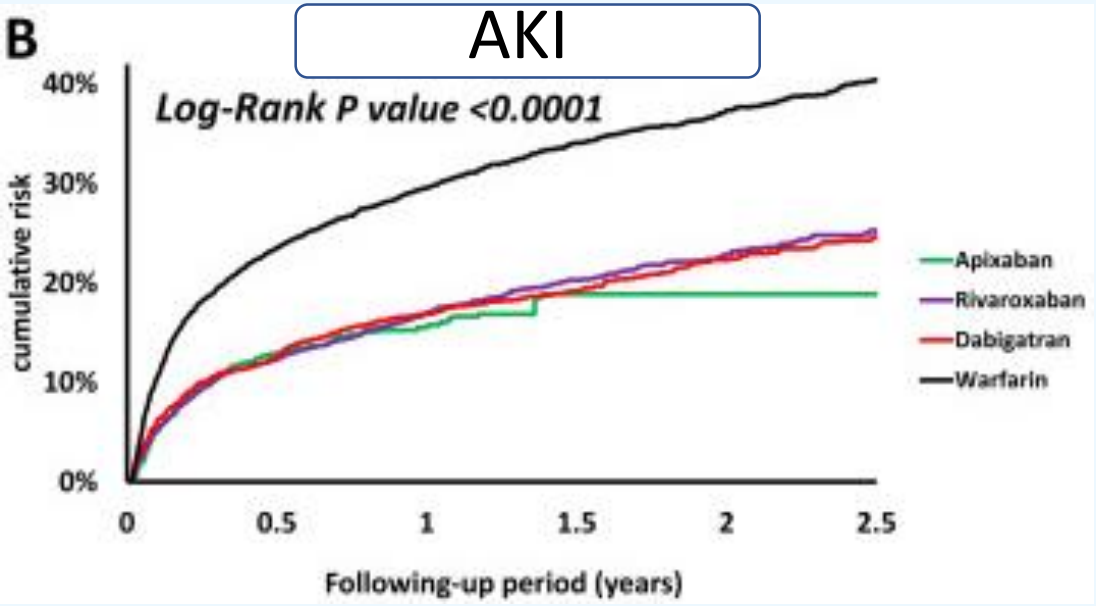
```
tebalance density mage, bwidth(*1.5)
```

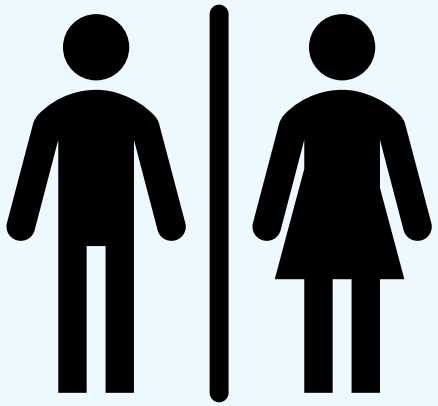


It has not been implemented for multivalued treatments

MULTIVALUED TREATMENTS

AKI





$$d = \frac{\alpha_1(1) + \alpha_3 Z}{\sigma}$$, hypothesis H_0: $\alpha_0 = \alpha_3 = 0$

1. Estimate PS from tx model
2. Logit sex tx ps tx##ps

Warfarin
Apixaban
Rivaroxaban
Dabigatran

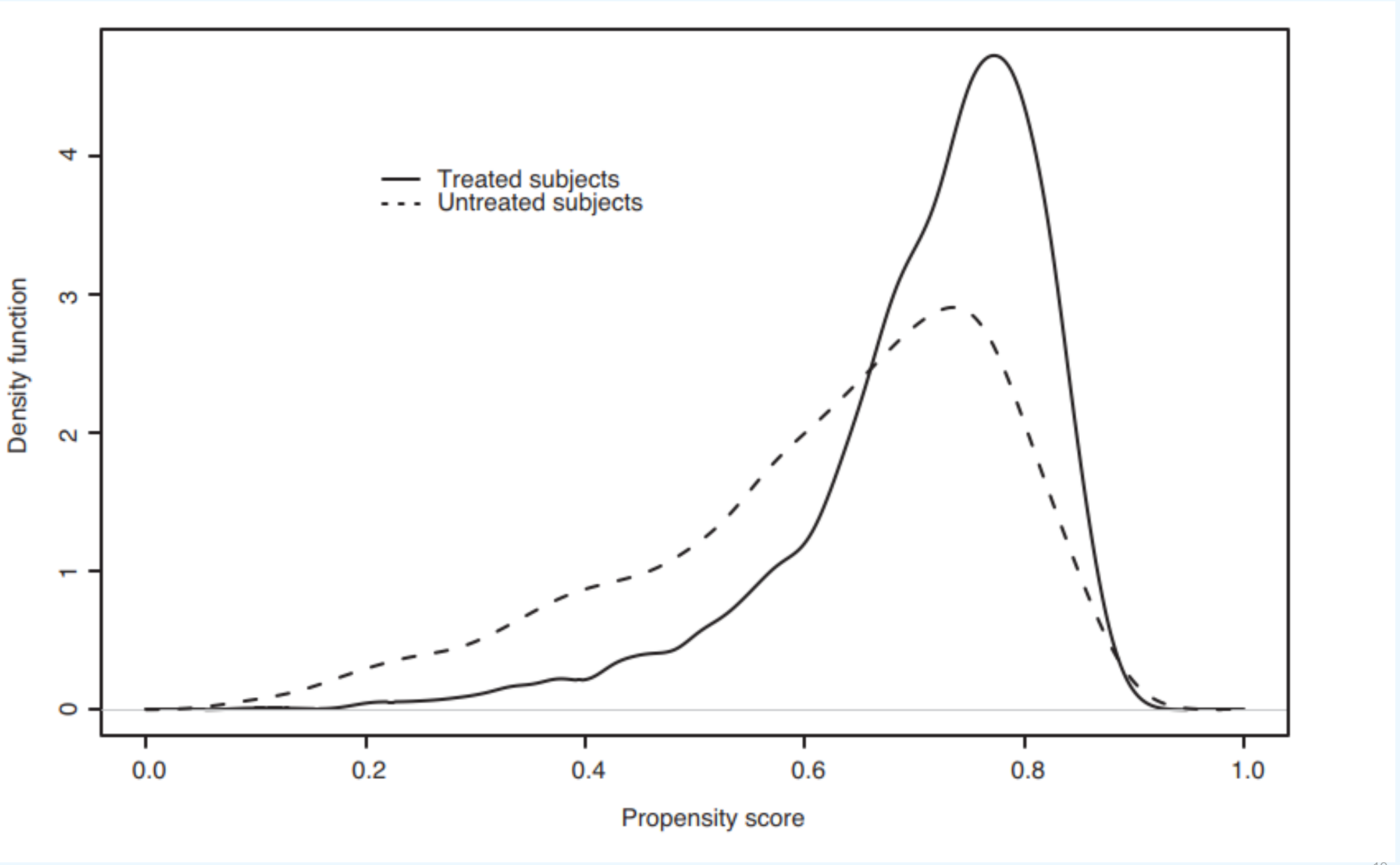
3. Test hypothesis $\_b_{warfarin} = \_b_{warfarin} * ps = 0$
   $\_b_{Rivaroxaban} = \_b_{Rivaroxaban} * ps = 0$
   $\_b_{Apixaban} = \_b_{Apixaban} * ps = 0$
   $\_b_{Dabigatran} = \_b_{Dabigatran} * ps = 0$

Limitation of testing for multivalued treatments

- There are limitations to balance tests based on statistical tests of Hypotheses
- First, the power of such tests is influenced by sample size. Therefore, given the same quantitative degree of balance between treated and untreated subjects, imbalance is more likely to be detected in larger samples than in smaller samples
- Second, in a given sample, the power of the above test may differ from that of a simple t-test comparing means between treated and untreated subjects in the overall sample.

Table 1. Baseline characteristics of treated and untreated subjects

Variable	Beta-blocker: No ($N=2929$)	Beta-blocker: Yes ($N=6178$)	p-value	Unconditional standardized difference	Weighted conditional standardized difference
Demographic characteristics					
Age	69.6 ± 13.5	65.0 ± 13.3	<.001	0.342	0.221
Female	1144 (39.1%)	1984 (32.1%)	<.001	0.147	0.087
Presenting signs and symptoms					
Acute CHF/Pulmonary edema	214 (7.3%)	224 (3.6%)	<.001	0.173	0.072
Classic cardiac risk factors					
Diabetes	842 (28.7%)	1494 (24.2%)	<.001	0.105	0.046
Current smoker	916 (31.3%)	2158 (34.9%)	<.001	0.077	0.041
CVA/TIA	354 (12.1%)	493 (8.0%)	<.001	0.142	0.115
Hyperlipidemia	767 (26.2%)	2132 (34.5%)	<.001	0.179	0.118
Hypertension	1343 (45.9%)	2793 (45.2%)	0.565	0.013	0.019
Family history of CAD	745 (25.4%)	2195 (35.5%)	<.001	0.217	0.105
Co-morbid conditions					
Angina	975 (33.3%)	1982 (32.1%)	0.251	0.026	0.044
Cancer	110 (3.8%)	154 (2.5%)	<.001	0.075	0.035
Dementia	142 (4.8%)	134 (2.2%)	<.001	0.157	0.063
Previous AMI	739 (25.2%)	1314 (21.3%)	<.001	0.095	0.045
Asthma	323 (11.0%)	181 (2.9%)	<.001	0.359	0.019
Depression	256 (8.7%)	377 (6.1%)	<.001	0.104	0.056
Peripheral vascular disease	281 (9.6%)	369 (6.0%)	<.001	0.141	0.068
Chronic CHF	189 (6.5%)	177 (2.9%)	<.001	0.183	0.099
Vital signs on admission					
Systolic BP	146.8 ± 31.4	149.9 ± 30.9	<.001	0.102	0.080
Diastolic BP	81.8 ± 18.6	84.9 ± 18.3	<.001	0.173	0.157
Heart rate	86.9 ± 25.9	82.1 ± 22.7	<.001	0.204	0.008
Respiratory rate	22.2 ± 6.5	20.3 ± 4.8	<.001	0.351	0.022
Laboratory tests					
Glucose	9.8 ± 5.2	9.2 ± 5.2	<.001	0.118	0.001
White blood count	10.6 ± 5.5	10.0 ± 4.3	<.001	0.130	0.010
Hemoglobin	135.2 ± 20.0	140.2 ± 17.7	<.001	0.267	0.134
Sodium	138.7 ± 4.2	139.2 ± 3.5	<.001	0.111	0.043
Potassium	4.1 ± 0.6	4.1 ± 0.5	<.001	0.127	0.038
Creatinine	114.2 ± 77.4	98.8 ± 50.3	<.001	0.254	0.008


Treated subjects
Untreated subjects
Density function
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Propensity score

Quantile regression to compare conditional distributions of continuous variables

Qualitatively compare

Quantile regression

Qualitatively compare the distribution of measured baseline covariates between treated and untreated subjects with the same propensity score.

Frequency the use of the 5th, 25th, 50th (median), 75th, and 95th regression quantiles.

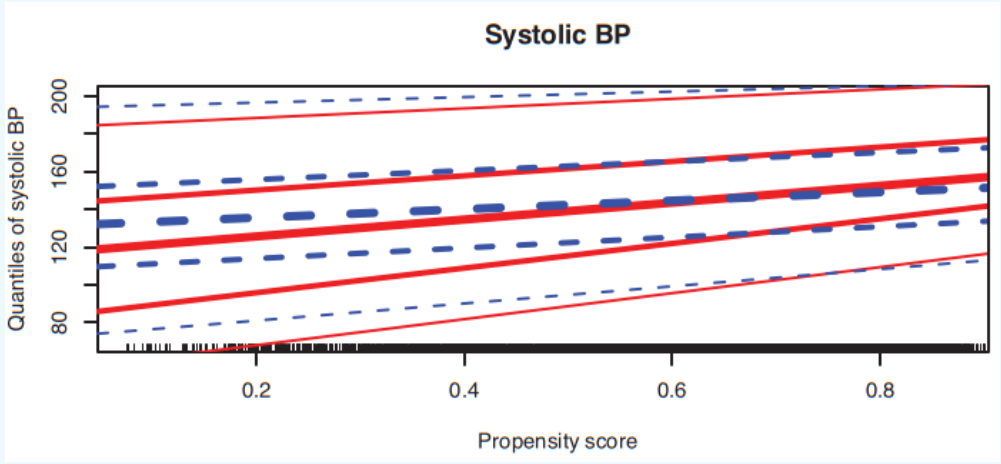
An advantage to the use of quantile regression compared is that one can examine how several quantiles (or percentiles) of the conditional response distribution vary with the predictor variables

$$\rho(X,\rho|T, Z) = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 T + \boldsymbol{\alpha}_2 Z + \boldsymbol{\alpha}_3 TxZ$$

$\rho(X,\rho|T, Z)$ denotes the conditional ρ th quantile of the covariate X

Quantile regression

Different regression quantiles.



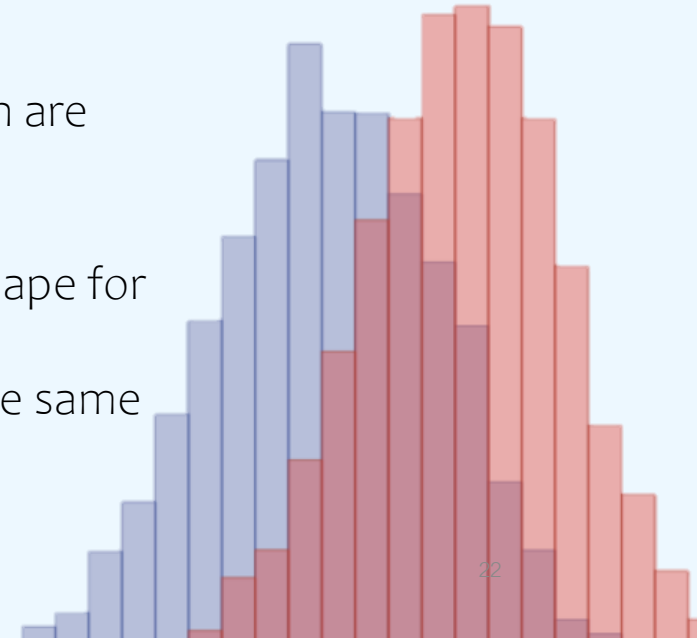
The use of quantile regression allows one to qualitatively examine the conditional distribution of a continuous baseline covariate.

Plotting the estimated regression quantiles against the estimated propensity score for treated and untreated subjects separately allows one to examine the distribution of X at specific values of Z, the estimated propensity score, in treated and untreated subjects.

The use of weighted conditional standardized absolute differences and quantile regression are intended to be complementary if two conditions are satisfied.

First, if conditional on the propensity score, the distribution of a baseline characteristic is symmetric within each treatment group and if the conditional distribution has the same shape for both treated and untreated.

Second, for treatment group, the conditional distribution of the baseline covariate is of the same shape for different values of the propensity score (i.e., only the location is shifted).



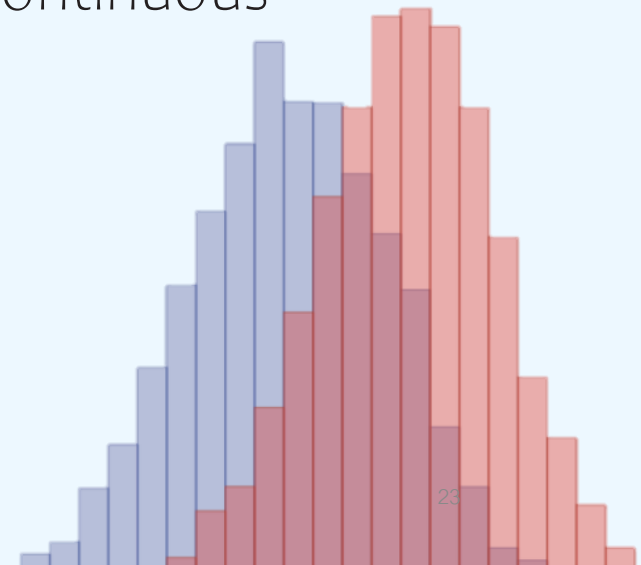
Comparison with balance diagnostics for stratification on the propensity score

Balance diagnostics : the distribution of baseline covariates between Tx and un-Tx within the same stratum (quintiles of the PS).

Two proposed of examining : within-quintile side-by-side boxplots Tx and un-Tx

1. Comparison of means between Tx and un-Tx within the same quintile.

2. Allows for a qualitative comparison of the distribution of continuous covariates between Tx and un-Tx within the same quintile.



CASE STUDY

We used data on 9107 patients who were discharged alive with an acute myocardial infarction (AMI or heart attack) from 102 hospitals in Ontario, Canada, between 1 April 1999 and 31 March 2001

The exposure of interest was whether the patient was prescribed a beta-blocker at hospital discharge, and the outcome of interest was death within 3 years of hospital discharge.

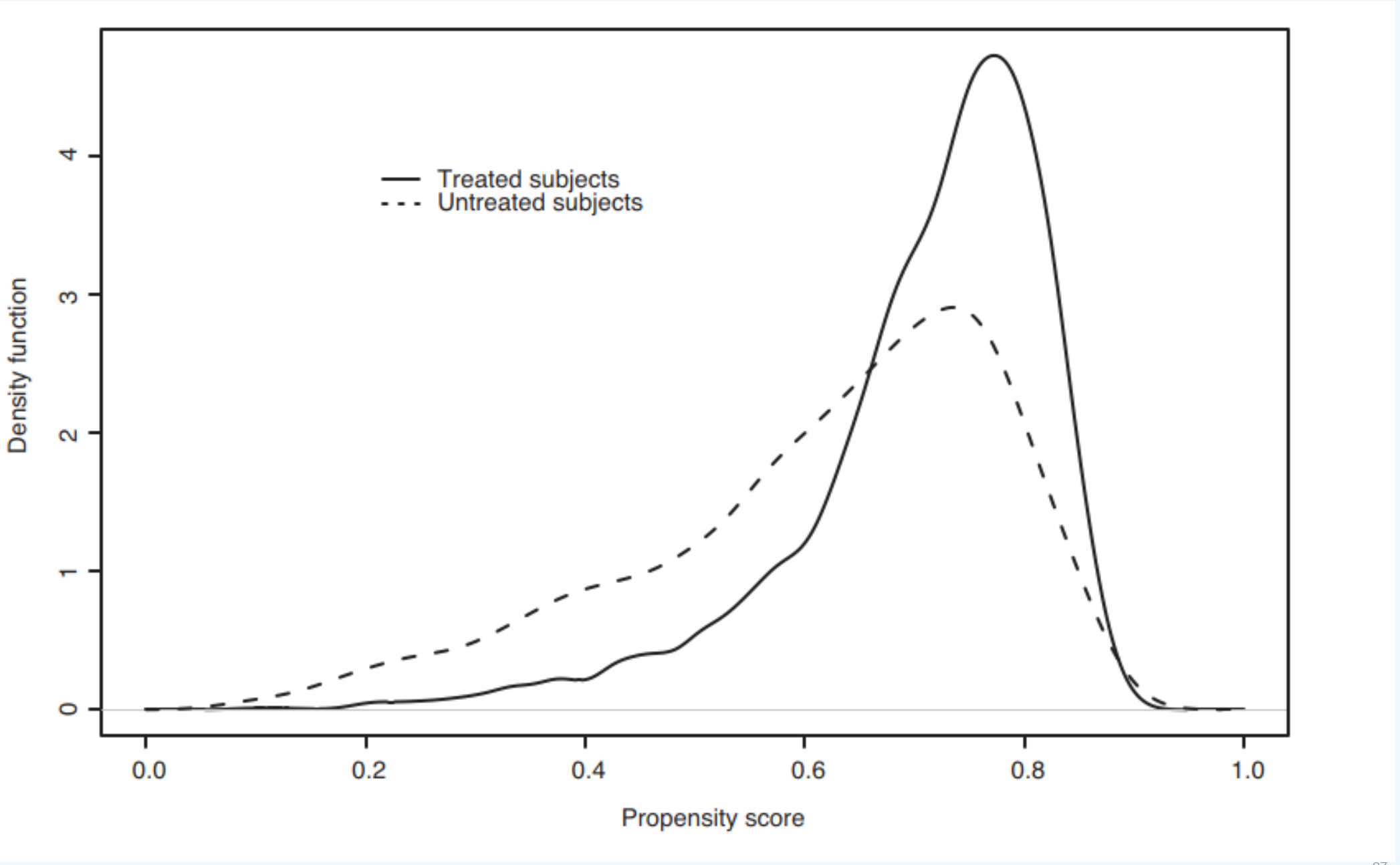
Overall, 6178 (67.8%) of patients received a beta-blocker at discharge, while 2929 (32.2%) did not receive a prescription at discharge.

PROPENSITY SCORE diagnostics

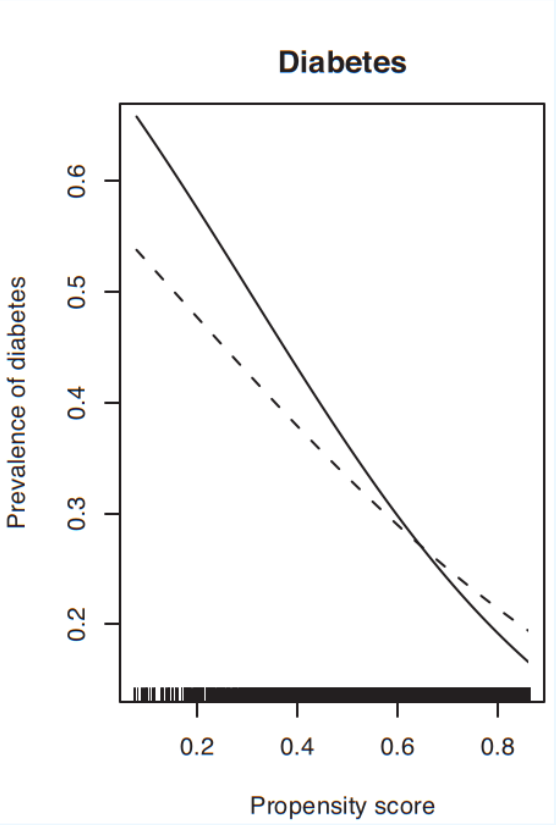
A propensity score model was fit using a logistic regression model in which treatment assignment (beta-blocker vs. no beta-blocker) was regressed on the 27 covariates listed in Table 1

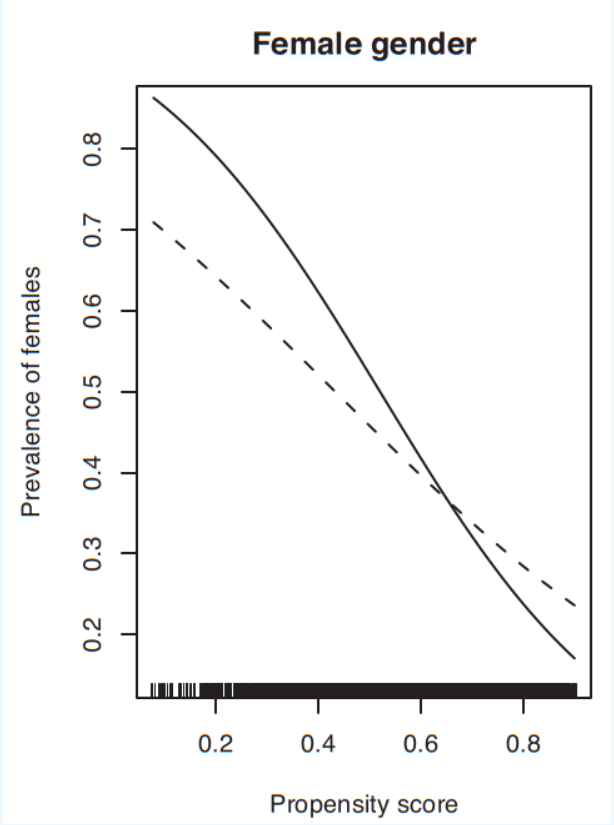
Table 1. Baseline characteristics of treated and untreated subjects

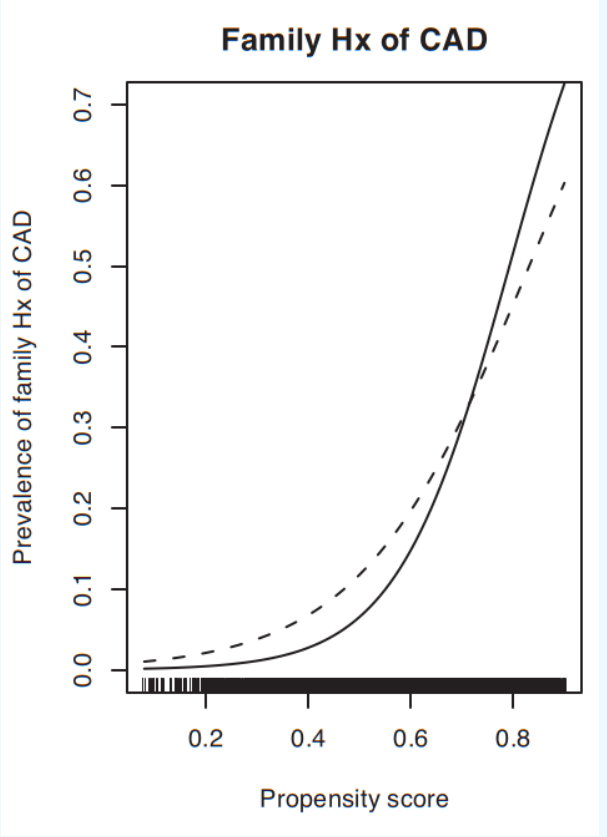
Variable	Beta-blocker: No ($N=2929$)	Beta-blocker: Yes ($N=6178$)	p-value	Unconditional standardized difference	Weighted conditional standardized difference
Demographic characteristics					
Age	69.6 ± 13.5	65.0 ± 13.3	<.001	0.342	0.221
Female	1144 (39.1%)	1984 (32.1%)	<.001	0.147	0.087
Presenting signs and symptoms					
Acute CHF/Pulmonary edema	214 (7.3%)	224 (3.6%)	<.001	0.173	0.072
Classic cardiac risk factors					
Diabetes	842 (28.7%)	1494 (24.2%)	<.001	0.105	0.046
Current smoker	916 (31.3%)	2158 (34.9%)	<.001	0.077	0.041
CVA/TIA	354 (12.1%)	493 (8.0%)	<.001	0.142	0.115
Hyperlipidemia	767 (26.2%)	2132 (34.5%)	<.001	0.179	0.118
Hypertension	1343 (45.9%)	2793 (45.2%)	0.565	0.013	0.019
Family history of CAD	745 (25.4%)	2195 (35.5%)	<.001	0.217	0.105
Co-morbid conditions					
Angina	975 (33.3%)	1982 (32.1%)	0.251	0.026	0.044
Cancer	110 (3.8%)	154 (2.5%)	<.001	0.075	0.035
Dementia	142 (4.8%)	134 (2.2%)	<.001	0.157	0.063
Previous AMI	739 (25.2%)	1314 (21.3%)	<.001	0.095	0.045
Asthma	323 (11.0%)	181 (2.9%)	<.001	0.359	0.019
Depression	256 (8.7%)	377 (6.1%)	<.001	0.104	0.056
Peripheral vascular disease	281 (9.6%)	369 (6.0%)	<.001	0.141	0.068
Chronic CHF	189 (6.5%)	177 (2.9%)	<.001	0.183	0.099
Vital signs on admission					
Systolic BP	146.8 ± 31.4	149.9 ± 30.9	<.001	0.102	0.080
Diastolic BP	81.8 ± 18.6	84.9 ± 18.3	<.001	0.173	0.157
Heart rate	86.9 ± 25.9	82.1 ± 22.7	<.001	0.204	0.008
Respiratory rate	22.2 ± 6.5	20.3 ± 4.8	<.001	0.351	0.022
Laboratory tests					
Glucose	9.8 ± 5.2	9.2 ± 5.2	<.001	0.118	0.001
White blood count	10.6 ± 5.5	10.0 ± 4.3	<.001	0.130	0.010
Hemoglobin	135.2 ± 20.0	140.2 ± 17.7	<.001	0.267	0.134
Sodium	138.7 ± 4.2	139.2 ± 3.5	<.001	0.111	0.043
Potassium	4.1 ± 0.6	4.1 ± 0.5	<.001	0.127	0.038
Creatinine	114.2 ± 77.4	98.8 ± 50.3	<.001	0.254	0.008

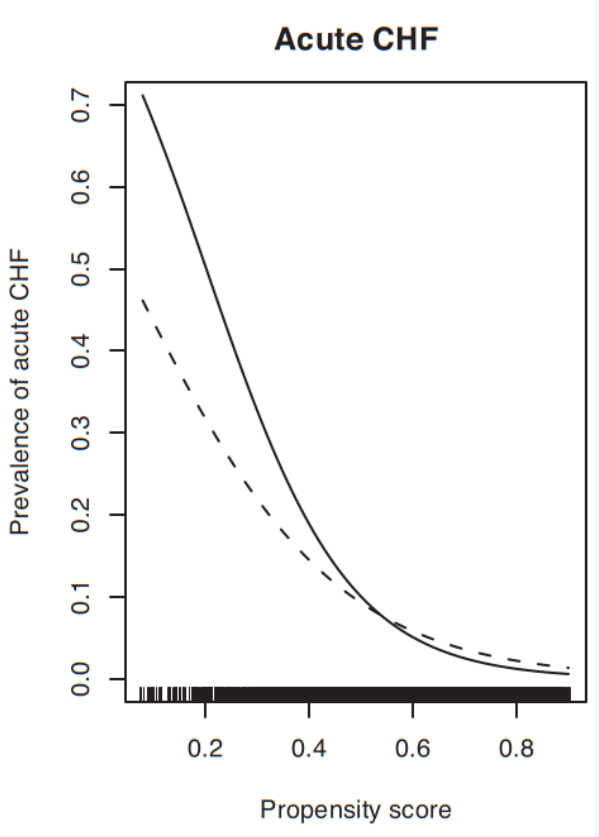

Treated subjects
Untreated subjects
Density function
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Propensity score

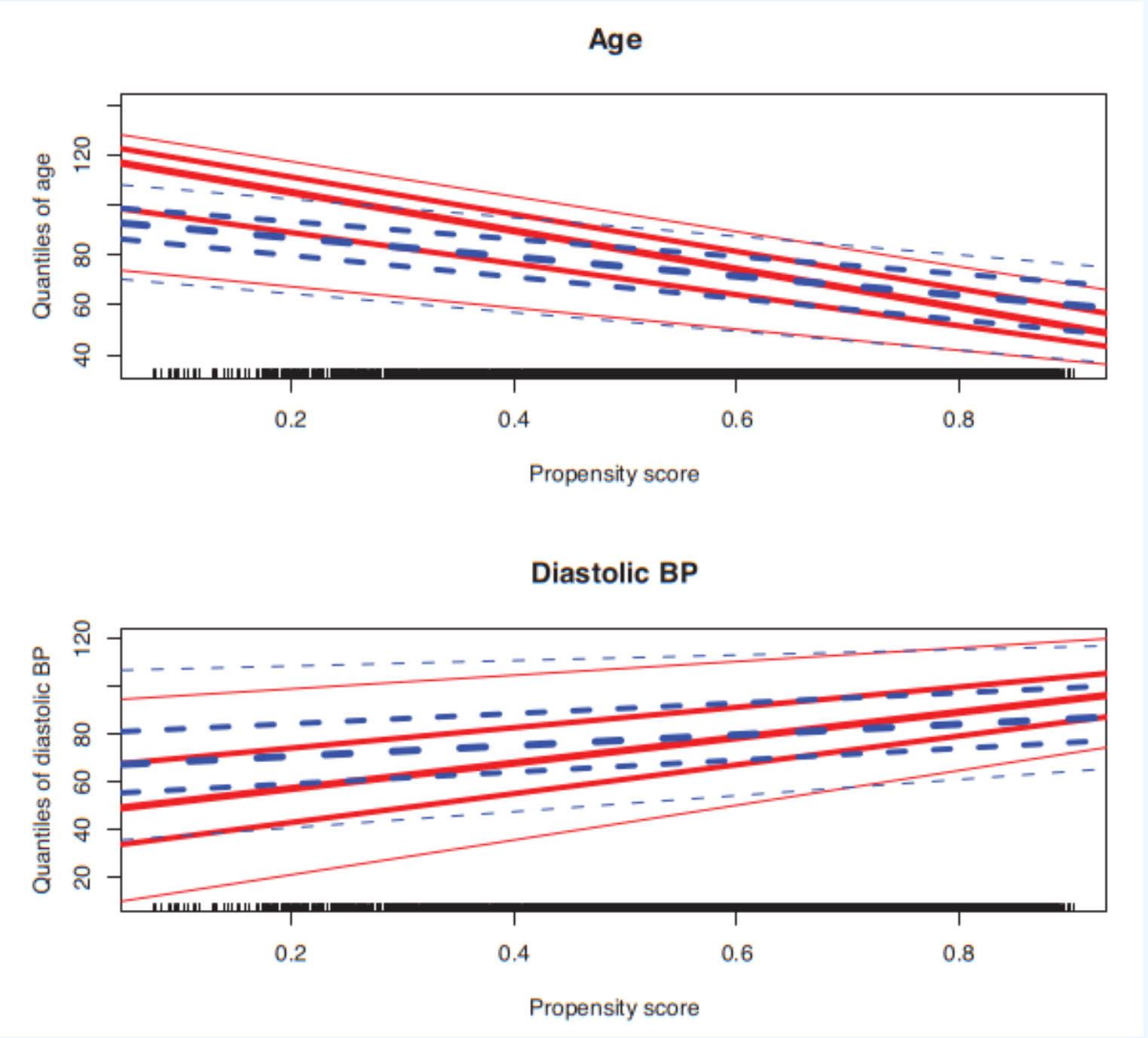
The prevalence of each dichotomous variables, conditional on the propensity score, in treated and untreated

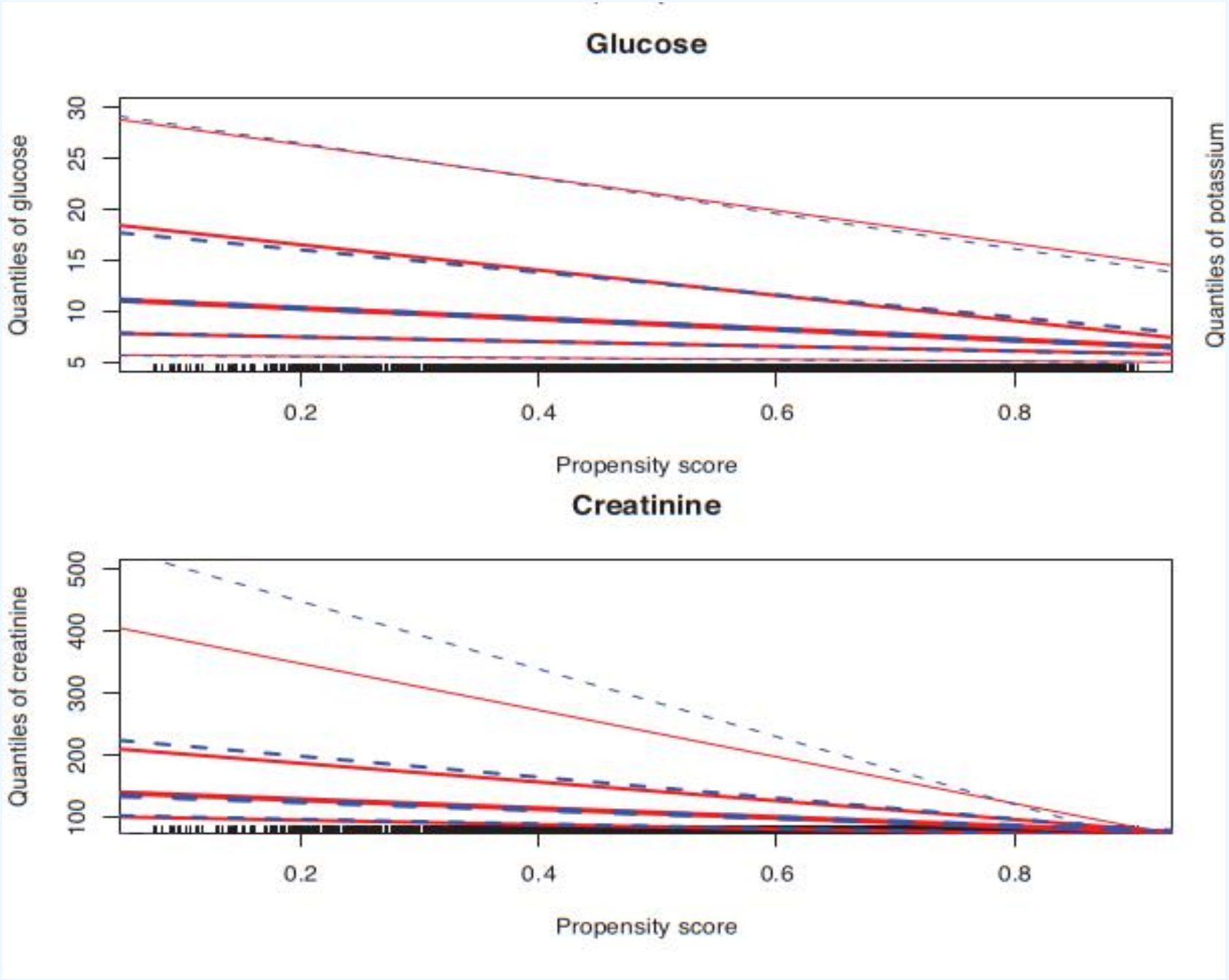









Age
Quantiles of age
40
60
80
120
0.2
0.4
0.6
0.8
Propensity score
Diastolic BP
Quantiles of diastolic BP
20
40
60
80
120
0.2
0.4
0.6
0.8
Propensity score


Glucose
Quantiles of glucose
5
10
15
20
25
30
0.2
0.4
0.6
0.8
Propensity score
Quantiles of potassium
Creatinine
Quantiles of creatinine
100
200
300
400
500
0.2
0.4
0.6
0.8
Propensity score

Discussion

- The focus of the current manuscript is balance diagnostics for when covariate adjustment using the propensity score is employed

- The developed the weighted conditional standardized absolute difference to quantitatively compare the conditional difference in baseline covariates between treated and untreated subjects.

- Quantile regression models be used to qualitatively examine the conditional distribution of continuous baseline covariates between treated and untreated subjects.

- However, no comparable methods have been proposed for when the propensity score is used for covariate adjustment


BALNACE