Propensity Score Diagnostics

Emily Granger

Jamie C. Sergeant

Mark Lunt

MANCHESTER 1824
The University of Manchester

CENTRE FOR EPIDEMIOLOGY VERSUS ARTHRITIS

@EGranger90

MRC Medical Research Council

Propensity scores are becoming increasingly popular

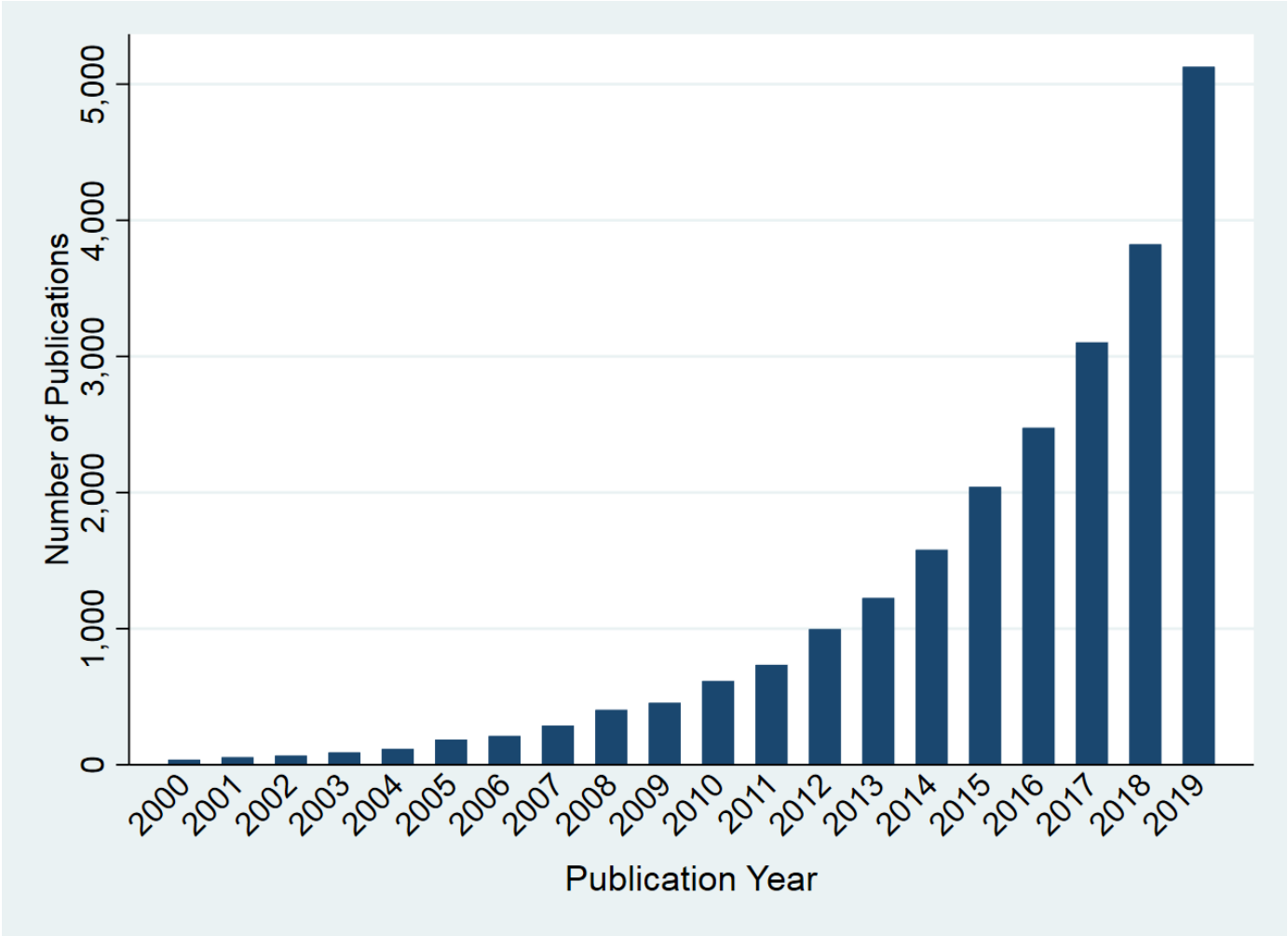


Figure 1: Number of propensity score publications in medical research by year

Review on the use of propensity score diagnostics

- Recent review on the use of propensity score diagnostics in the applied medical literature [Granger et al. 2020]

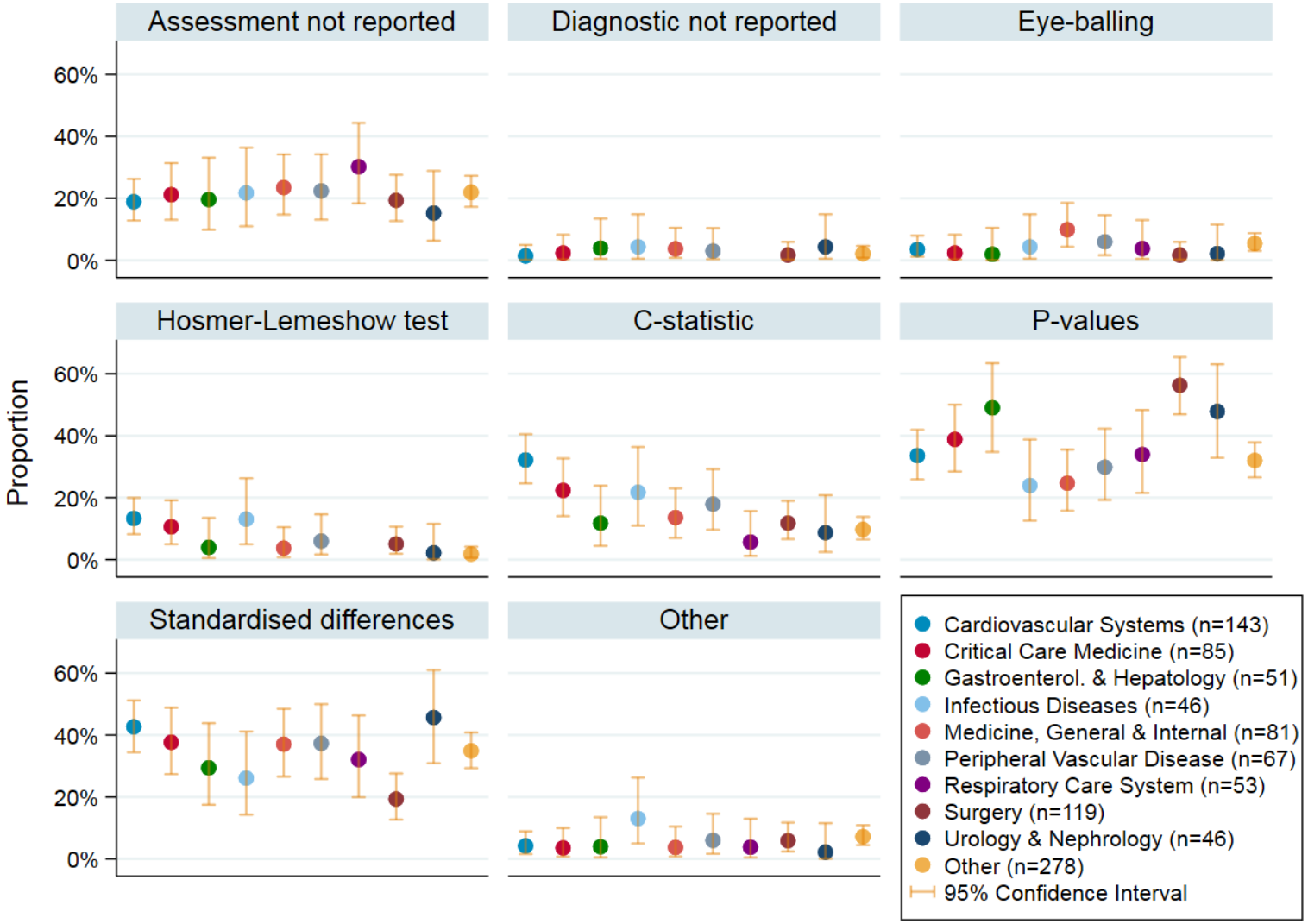
- Inclusion criteria:
  - Publication years 2014-2016
  - High-impact journals (Impact Factor > 4)

- Extracted data on:
  - Research area
  - Propensity score method used
  - Diagnostics used

Review on the use of propensity score diagnostics

Key Findings:

- 894 studies included
- 20.9% did not report use of any diagnostic
- 36.6% used hypothesis tests



Aims of research

Aim 1:

Review and compare the existing propensity score diagnostics.

Aim 2:

Develop guidelines for how to build and assess propensity score models.

Aims of research

Individual Diagnostics

Overall Diagnostics

Aim 2:

Develop guidelines for how to build and assess propensity score models.

Aims of research

Individual Diagnostics

Overall Diagnostics

Aim 2:

Develop guidelines for how to build and assess propensity score models.

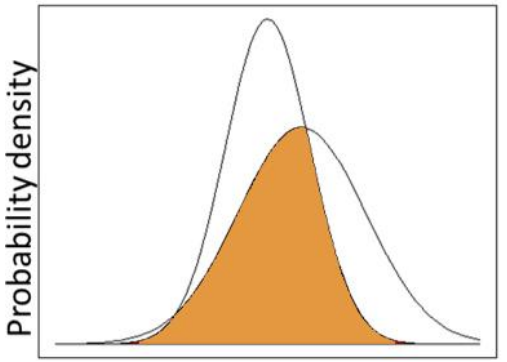
Individual diagnostics

Mean-based

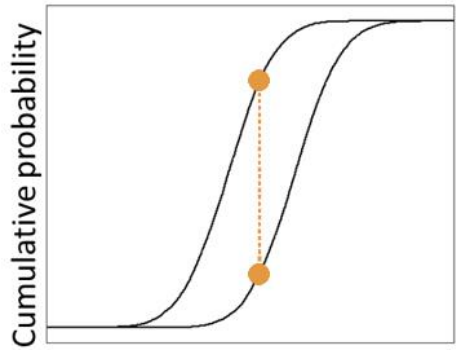
- Standardised difference (SD)
- t-test statistic (t)
- Percent reduction in mean difference (PR)

Distribution-based

- Overlapping coefficient (OVL)



- Kolmogorov-Smirnov Statistic (KS)

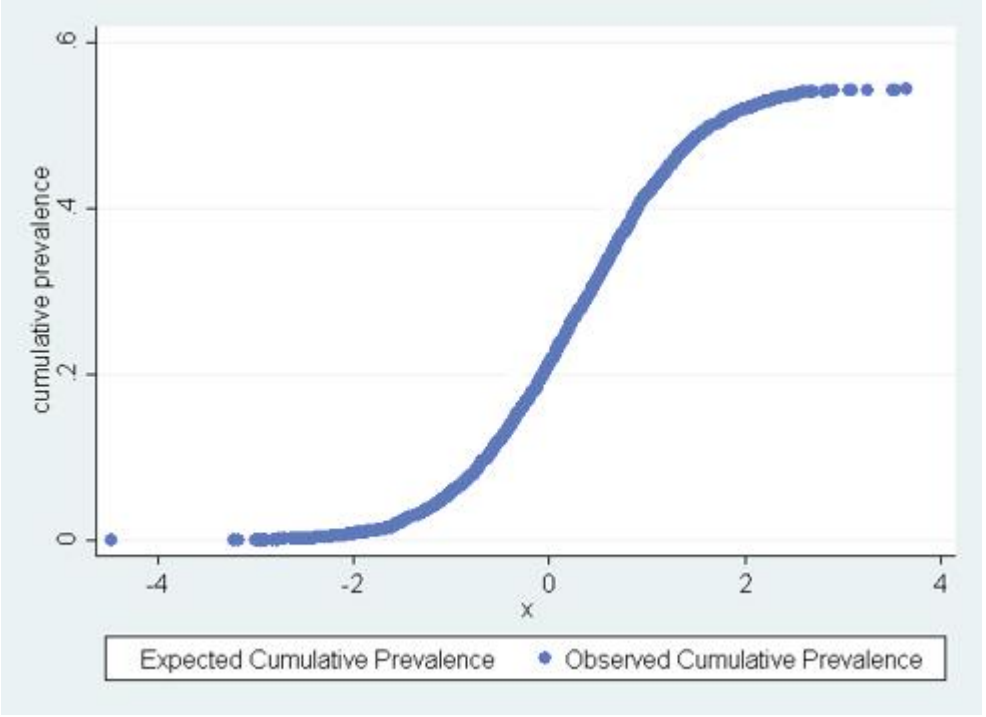


Cumulative prevalence of exposure

Notation: exposure indicator for subject i: E_i, propensity score for subject i: PS_i, sample size: n.

For continuous variable X:

- $OCP_X(X_0) = \frac{1}{n}\sum_{i:X_i \leq X_0} E_i$

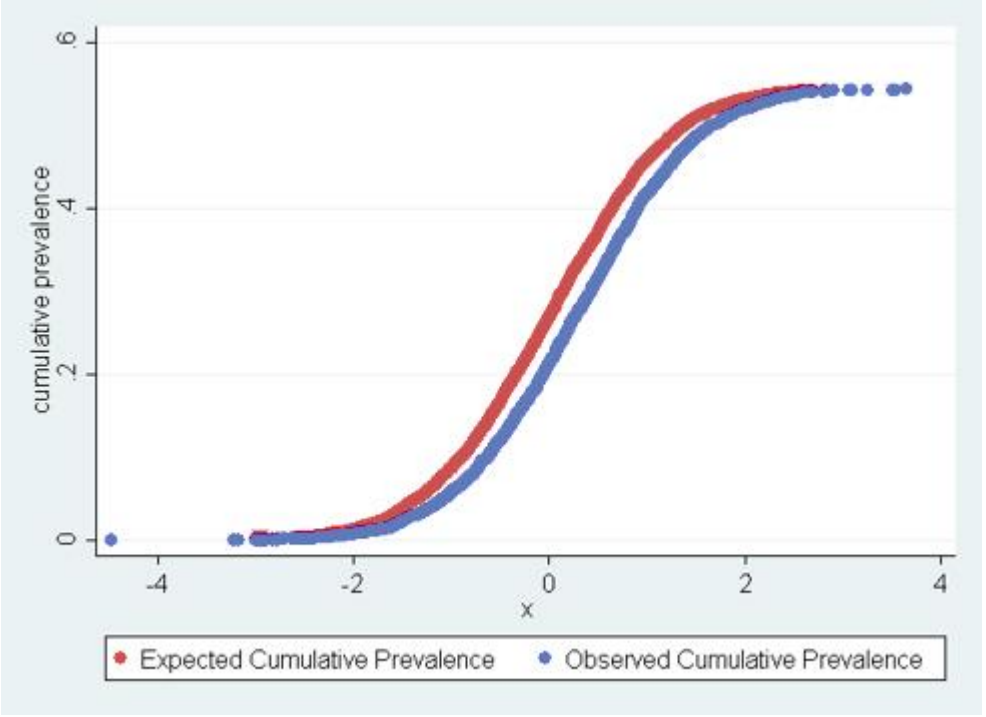


Cumulative prevalence of exposure

Notation: exposure indicator for subject i: E_i, propensity score for subject i: PS_i, sample size: n.

For continuous variable X:

- $OCP_X(X_0) = \frac{1}{n}\sum_{i:X_i \leq X_0} E_i$
- $ECP_X(X_0) = \frac{1}{n}\sum_{i:X_i \leq X_0} PS_i$

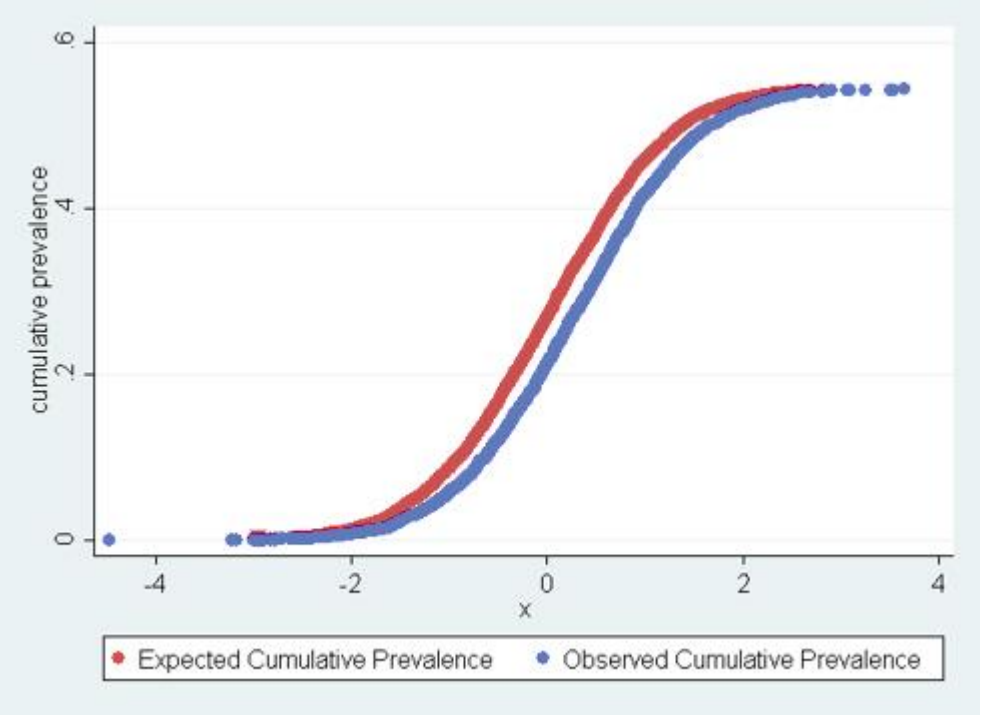


Cumulative prevalence of exposure

Notation: exposure indicator for subject i: E_i, propensity score for subject i: PS_i, sample size: n.

For continuous variable X:

- $OCP_X(X_0) = \frac{1}{n}\sum_{i:X_i \leq X_0} E_i$
- $ECP_X(X_0) = \frac{1}{n}\sum_{i:X_i \leq X_0} PS_i$
- $D_X = |\, OCP_X - ECP_X \,|$



Simulated data

Propensity score model:

- logit(PS)=$\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 +, \ldots, \alpha_7 X_7 + \alpha_8 X_8$

Variation between scenarios:

Correct PS:

S1: $X_8 = 0$	Linear model

Simulated data

Propensity score model:

- logit(PS)=$\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 +, \ldots, \alpha_7 X_7 + \alpha_8 X_8$

Variation between scenarios:

Correct PS:

S1: $X_8 = 0$	Linear model
S2: $X_8 = 0.4(3.5^{X_1} - 1)$	**Nonlinearity added (monotonic)**

Simulated data

Propensity score model:

- logit(PS)=$\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 +, \ldots, \alpha_7 X_7 + \alpha_8 X_8$

Variation between scenarios:

Correct PS:

S1: $X_8 = 0$	Linear model
S2: $X_8 = 0.4(3.5^{X_1} - 1)$	Nonlinearity added (monotonic)
S3: $\boldsymbol{X_8 = X_4 X_5}$	**Binary-binary interaction**
S4: $\boldsymbol{X_8 = X_4 X_1}$	**Binary-continuous interaction**
S5: $\boldsymbol{X_8 = X_1 X_2}$	**Continuous-continuous interaction**

Simulated data

Propensity score model:

- logit(PS)=$\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 +, \ldots, \alpha_7 X_7 + \alpha_8 X_8$

Variation between scenarios:

Correct PS:		Incorrect PS:
S1: $X_8 = 0$	Linear model	$X_1 = 0$
S2: $X_8 = 0.4(3.5^{X_1} - 1)$	Nonlinearity added (monotonic)	$X_8 = 0$
S3: $\boldsymbol{X_8 = X_4 X_5}$	**Binary-binary interaction**	$\boldsymbol{X_8 = 0}$
S4: $\boldsymbol{X_8 = X_4 X_1}$	**Binary-continuous interaction**	$\boldsymbol{X_8 = 0}$
S5: $\boldsymbol{X_8 = X_1 X_2}$	**Continuous-continuous interaction**	$\boldsymbol{X_8 = 0}$

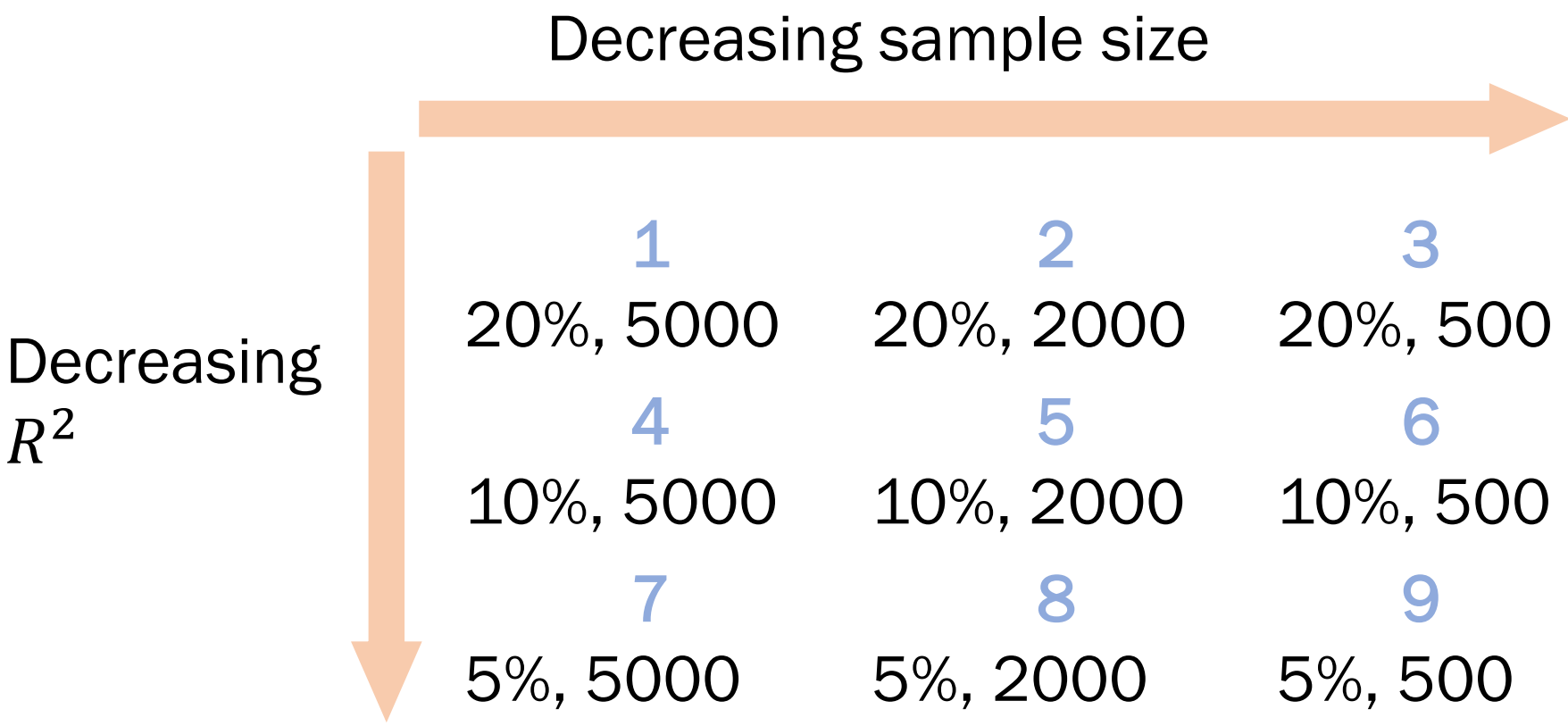
Scenario 1: Omission of a linear term

SD: standardised difference	KS: Kolmogorov-Smirnov statistic
t: t-test statistic	OVL: overlapping coefficient
PR: percent reduction in mean prevalence	CP: cumulative prevalence

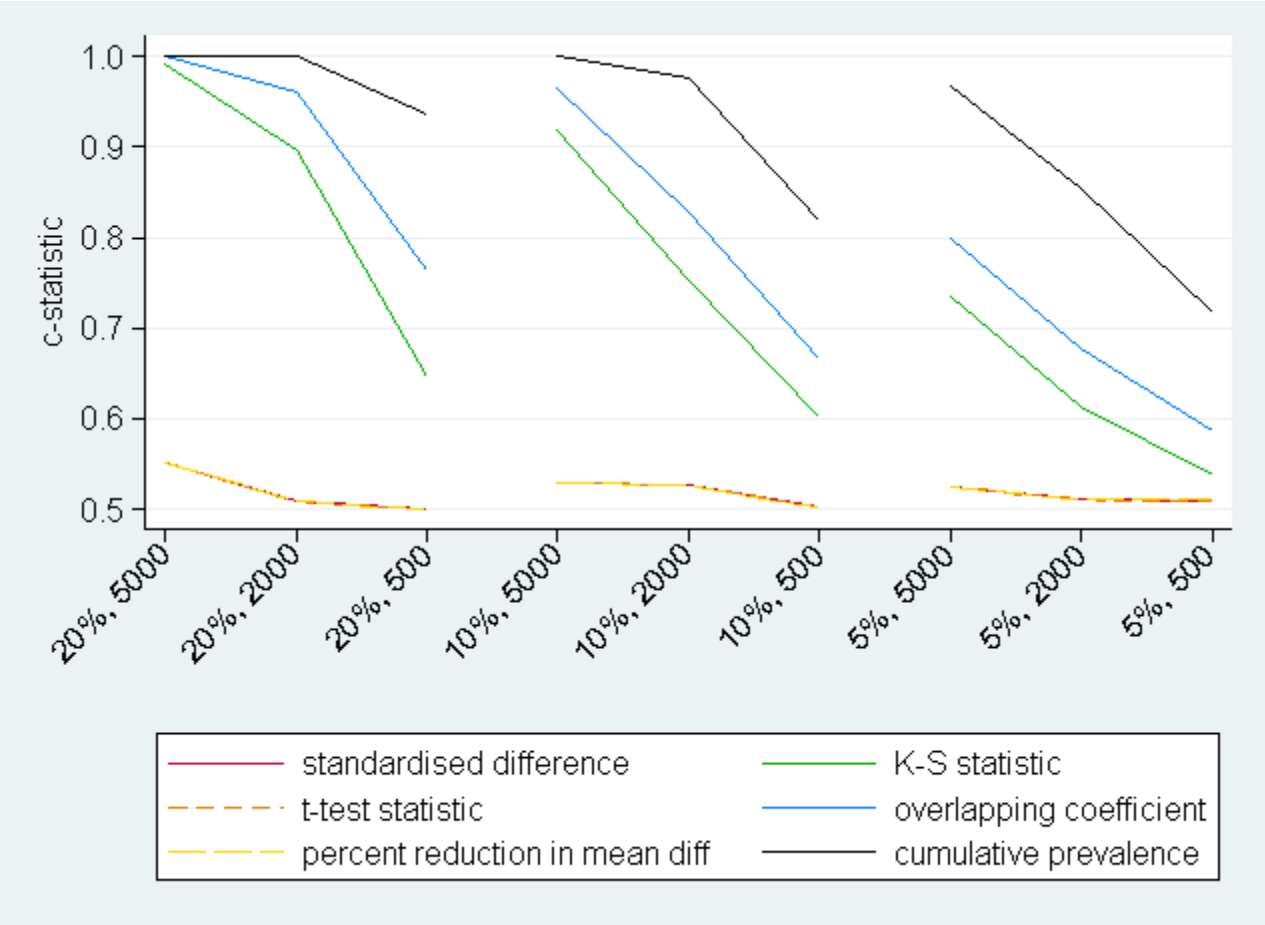
Scenario 1: Omission of a linear term

SD: standardised difference	KS: Kolmogorov-Smirnov statistic
t: t-test statistic	OVL: overlapping coefficient
PR: percent reduction in mean prevalence	CP: cumulative prevalence

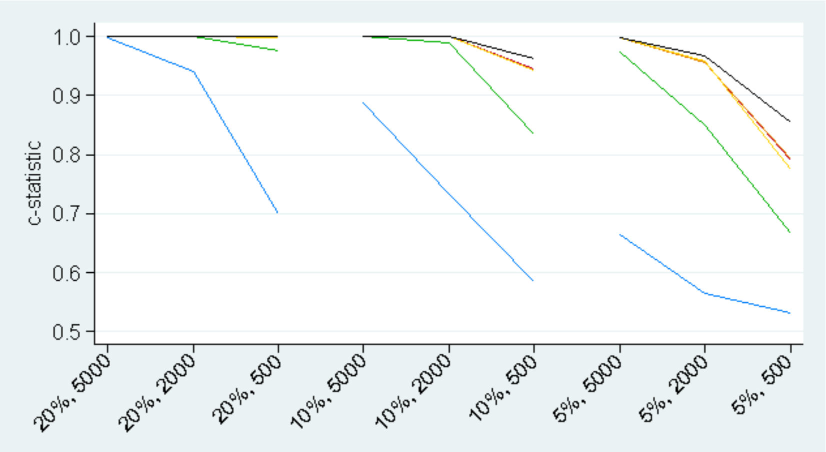
Simulated data

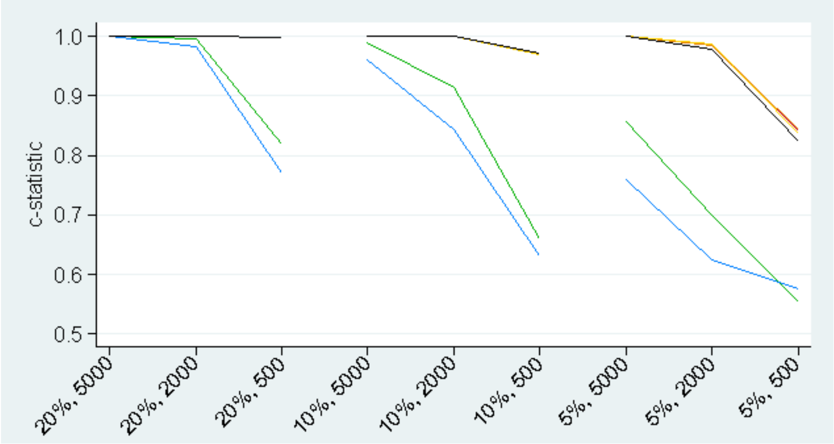


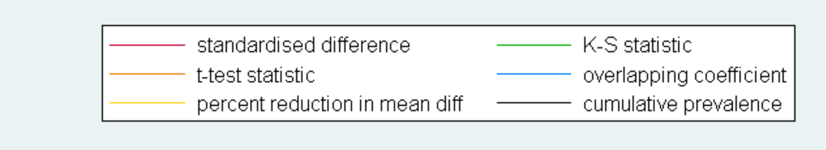
Scenario 2: Misspecification of a non-linear term



Scenarios 3-5: Omission of an interaction term







Figures:

- Scenario 3 (top left) binary-binary
- Scenario 4 (top right) binary-continuous
- Scenario 5 (bottom left) continuous-continuous

Conclusions (so far)

- Mean-based diagnostics can fail to identify nonlinear misspecifications in the propensity score
- Distribution-based diagnostics least reliable at identifying omission of interactions terms.
- Cumulative prevalence diagnostics most useful for identifying all types of propensity score misspecification.

But……

What about the outcome?

Published in final edited form as:
Am J Epidemiol. 2006 June 15; 163(12): 1149–1156.

Variable selection for propensity score models.

M. Alan Brookhart[1], **Sebastian Schneeweiss**[1], **Kenneth J. Rothman**[1,2], **Robert J. Glynn**[1,3], **Jerry Avorn**[1], and **Til Stürmer**[1]
1 *Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA*

Published in final edited form as:
Pharmacoepidemiol Drug Saf. 2011 June ; 20(6): 551–559. doi:10.1002/pds.2098.

The implications of propensity score variable selection strategies in pharmacoepidemiology – an empirical illustration

Amanda R. Patrick[1], **Sebastian Schneeweiss**[1], **M. Alan Brookhart**[2], **Robert J. Glynn**[1,3], **Kenneth J. Rothman**[4], **Jerry Avorn**[1], and **Til Stürmer**[2]

[1]Division of Pharmacoepidemiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

[2]Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, North Carolina

[3]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts

[4]RTI Health Solutions, Research Triangle Park, NC

Abstract

Purpose—To examine the effect of variable selection strategies on the performance of propensity score (PS) methods in a study of statin initiation, mortality and hip fracture assuming a true mortality reduction of <15% and no effect on hip fracture.

Methods—We compared seniors initiating statins with seniors initiating glaucoma medications. Out of 202 covariates with a prevalence > 5%, PS variable selection strategies included none, a priori, factors predicting exposure, and factors predicting outcome. We estimated hazard ratios

Aims of research

Individual Diagnostics

Overall Diagnostics

Aim 2:

Develop guidelines for how to build and assess propensity score models.

Overall diagnostics

Which balance metric?

- Standardised difference (SD)
- Overlapping coefficient (OVL)
- Kolmogorov-Smirnov Statistic (KS)

Which weighting scheme?

Let w_{ji} denote the j^{th} weight for covariate i. Then:

- $w_{1i} = \gamma_i Std.Dev(x_i)$ [Caruana et al. 2015]
  - $\gamma_i$is the coefficient for x_i obtained after regressing outcome on x_i.
- $w_{2i} = \delta_i Std.Dev(x_i)$
  - $\delta_i$is the coefficient for x_i obtained after regressing outcome on all covariates.

Overall diagnostics

- Disease risk scores (DRS) defined as predicted outcome under the control condition
- Standardised mean difference in DRS as a propensity score diagnostic [Stuart et al. 2013]

Simulated data

Propensity score model:

- logit(PS)=$\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 +, \ldots, \alpha_9 X_9$

Outcome model:

- Y=$\beta_0 + \beta_1 X_1 + \beta_2 X_2 +, \ldots, \beta_9 X_9 + \beta_{10} X_{10}$

Linear and Non-linear Scenarios:

S1: $X_{10} = 0$	Independent baseline covariates
S2: $X_{10} = 0$	Correlated baseline covariates
S3: $X_{10} = 0.2(6.0^{X_1} - 1)$	Monotonic non-linearity

Simulated data

Propensity score model:

- logit(PS)=$\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 +, \ldots, \alpha_9 X_9$

Outcome model:

- Y=$\beta_0 + \beta_1 X_1 + \beta_2 X_2 +, \ldots, \beta_9 X_9 + \beta_{10} X_{10}$

Non-additive Scenarios:

S4: $X_{10} = X_1 X_5$	Binary-binary interaction
S5: $X_{10} = X_1 X_2$	Binary-continuous interaction
S6: $X_{10} = X_2 X_7$	Continuous-continuous interaction

Scenarios 1 and 2: Linear outcomes

Table 1: Spearman rank correlation between overall diagnostics and bias

Scenario	Balance Metric	Weights 1	Weights 2	SD(DRS)
Scenario 1	SD	0.992	0.996	1.000
	KS	0.137	0.134	
	OVL	0.012	0.016	
Scenario 2	SD	0.129	0.995	1.000
	KS	0.102	0.142	
	OVL	0.031	0.053	

*SD: Standardised difference; KS: Kolmogorov-Smirnov statistic; OVL: Overlapping coefficient; DRS: Disease risk score

Scenarios 1 and 2: Linear outcomes

Table 1: Spearman rank correlation between overall diagnostics and bias

Scenario	Balance Metric	Weights 1	Weights 2	SD(DRS)
Scenario 1	SD	0.992	0.996	1.000
	KS	0.137	0.134	
	OVL	0.012	0.016	
Scenario 2	SD	0.129	0.995	1.000
	KS	0.102	0.142	
	OVL	0.031	0.053	

*SD: Standardised difference; KS: Kolmogorov-Smirnov statistic; OVL: Overlapping coefficient; DRS: Disease risk score

Scenarios 1 and 2: Linear outcomes

Table 1: Spearman rank correlation between overall diagnostics and bias

Scenario	Balance Metric	Weights 1	Weights 2	SD(DRS)
Scenario 1	SD	0.992	0.996	1.000
	KS	0.137	0.134	
	OVL	0.012	0.016	
Scenario 2	SD	0.129	0.995	1.000
	KS	0.102	0.142	
	OVL	0.031	0.053	

*SD: Standardised difference; KS: Kolmogorov-Smirnov statistic; OVL: Overlapping coefficient; DRS: Disease risk score

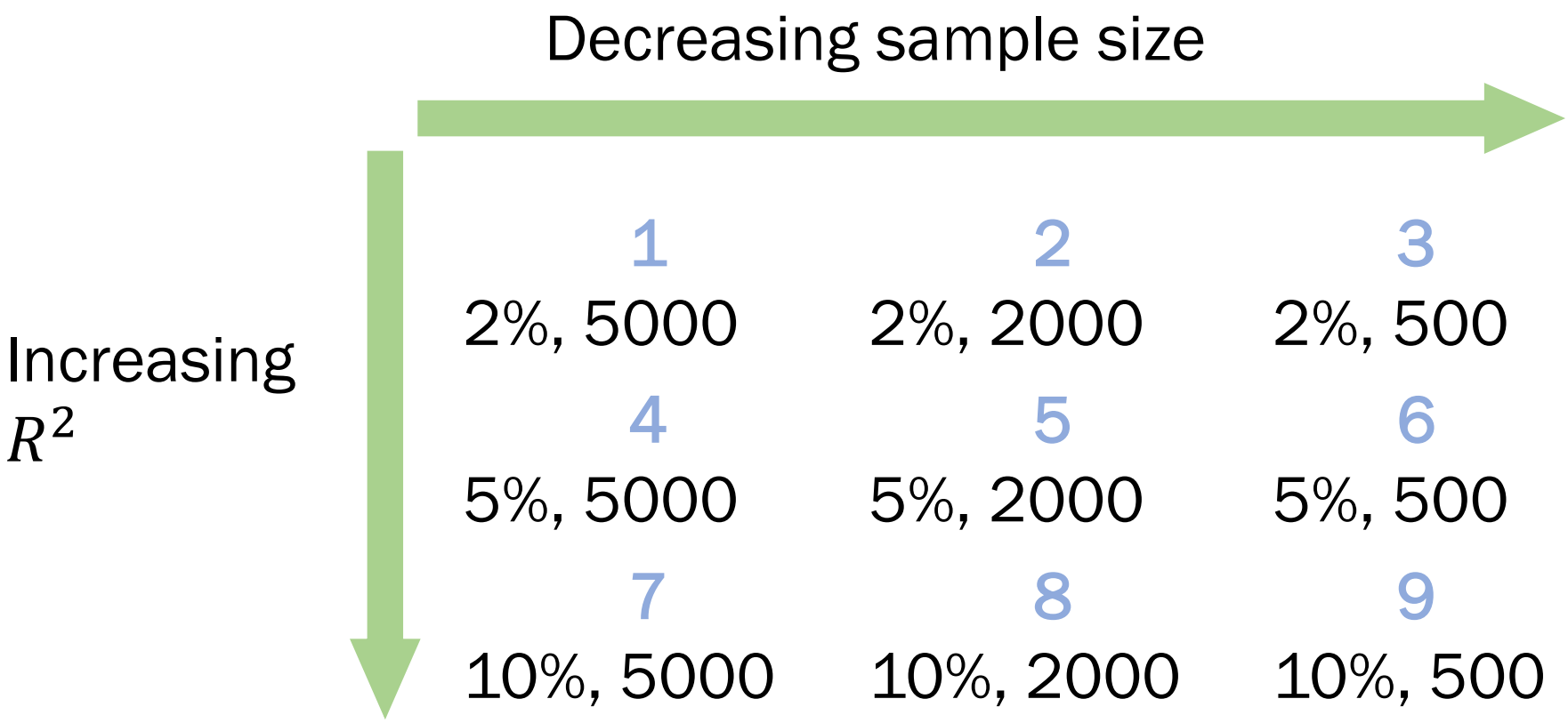
Scenarios 1 and 2: Linear outcomes

Table 1: Spearman rank correlation between overall diagnostics and bias

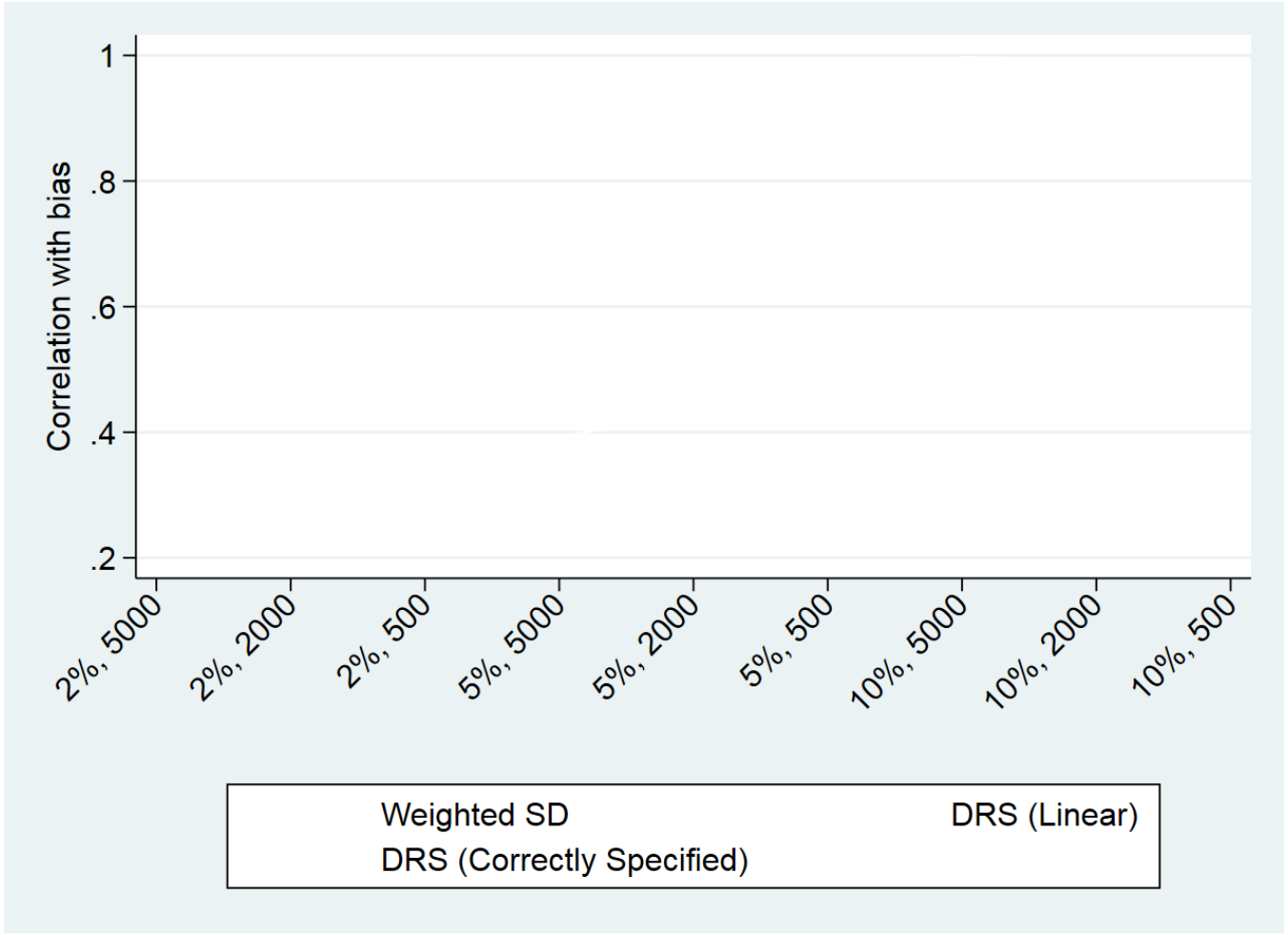
Scenario	Balance Metric	Weights 1	Weights 2	SD(DRS)
Scenario 1	SD	0.992	0.996	1.000
	KS	0.137	0.134	
	OVL	0.012	0.016	
Scenario 2	SD	0.129	0.995	1.000
	KS	0.102	0.142	
	OVL	0.031	0.053	

*SD: Standardised difference; KS: Kolmogorov-Smirnov statistic; OVL: Overlapping coefficient; DRS: Disease risk score

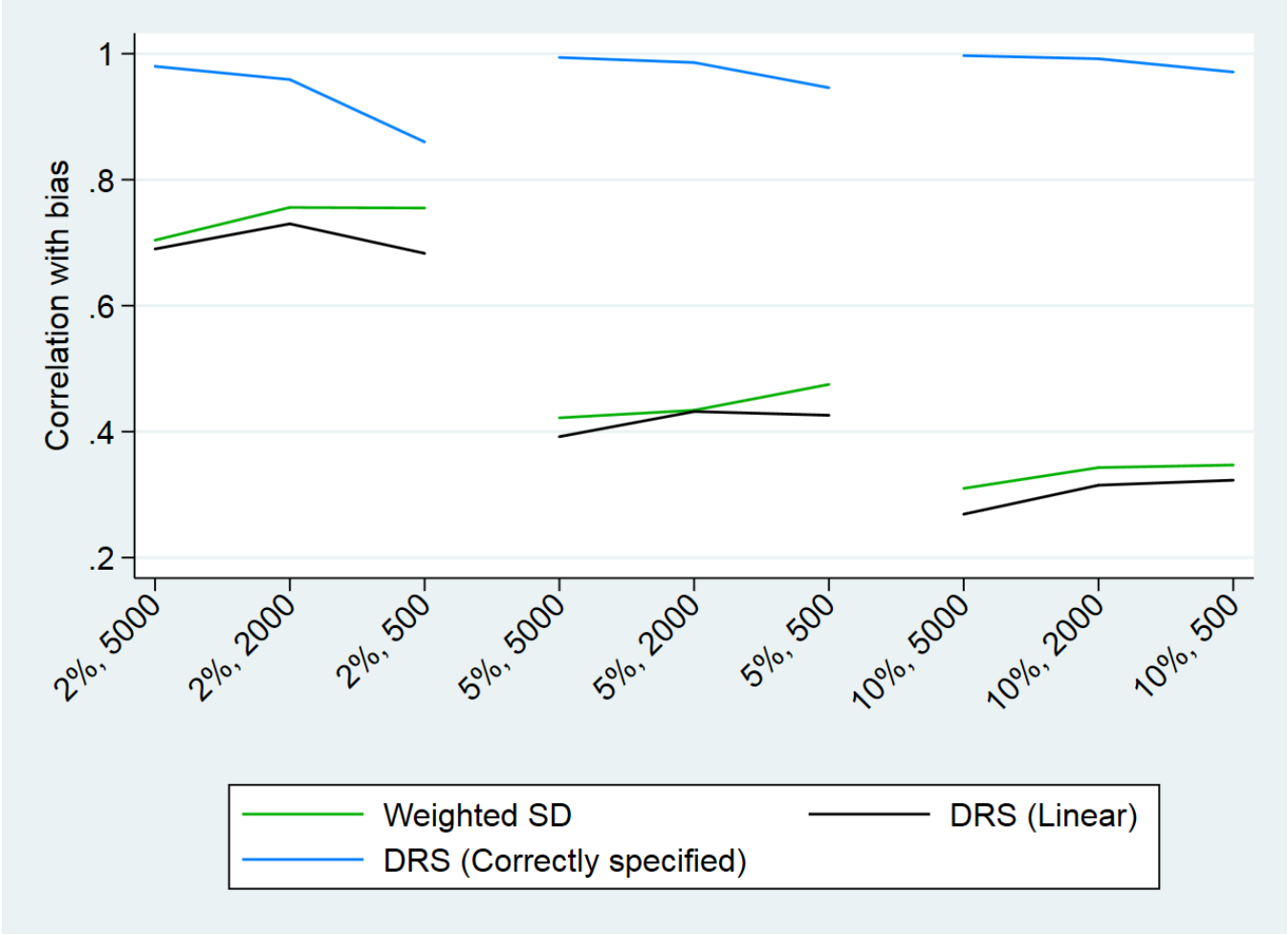
Simulated data



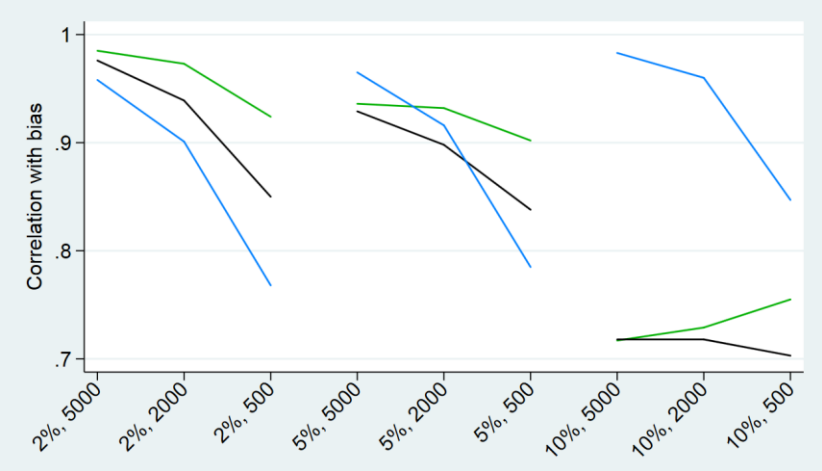
Scenario 2: Non-linear term in outcome model

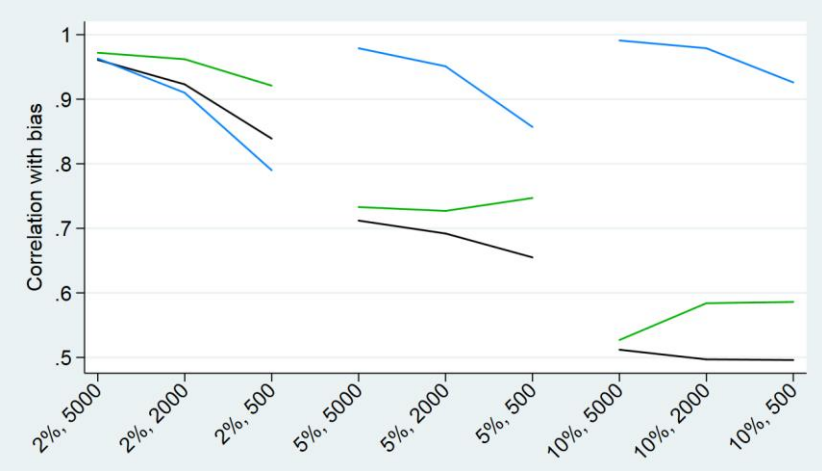


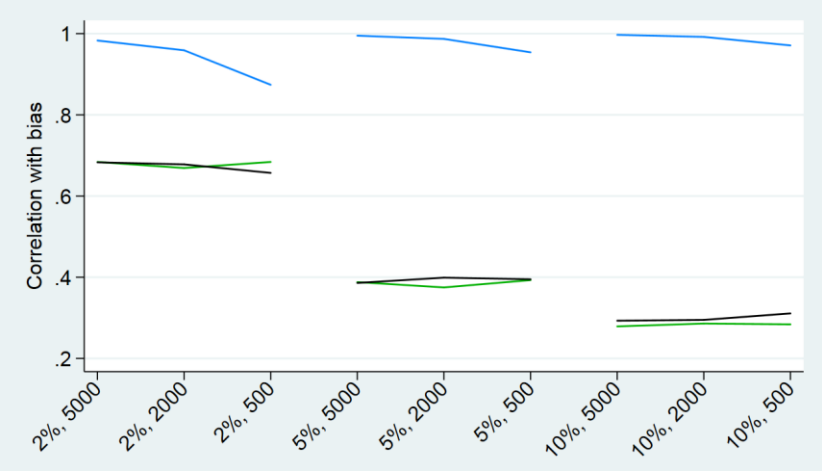
Scenario 2: Non-linear term in outcome model

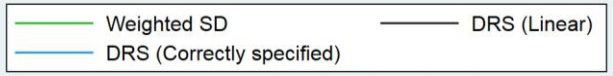


Scenarios 3-5: Interaction term in the outcome model









Figures:

- Scenario 3 (top left) binary-binary
- Scenario 4 (top right) binary-continuous
- Scenario 5 (bottom left) continuous-continuous

Conclusions

- Main finding: Standardised mean difference in the disease risk score is a promising overall diagnostic

- Limitations:
  (1) Not robust to misspecifications in the outcome model
  (2) Performance dependent on sample size

- Possible solutions:
  (1) Use of CP diagnostics to check specification
  (2) Using full sample or historic cohort to estimate DRS

Aims of research

Individual Diagnostics

Overall Diagnostics

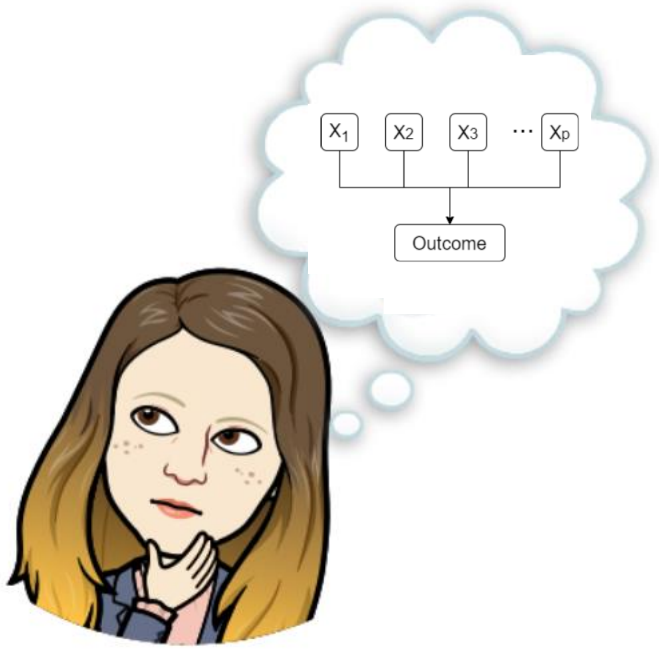
Aim 2:

Develop guidelines for how to build and assess propensity score models.

So, how best to assess propensity scores?

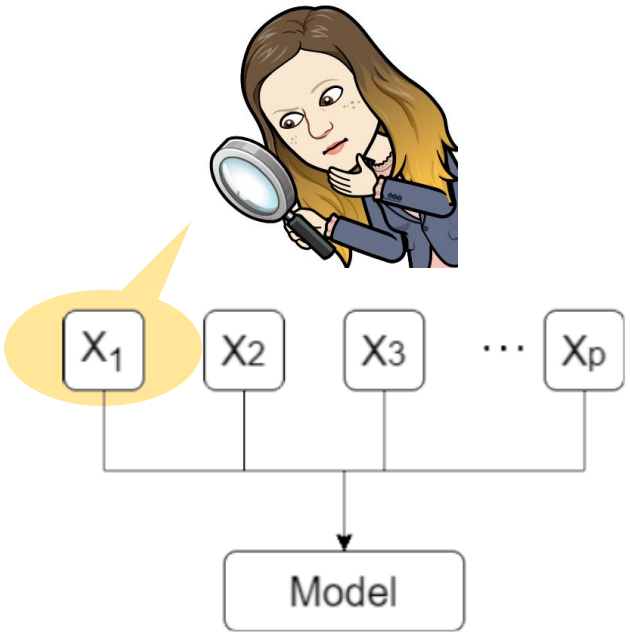
STEP 1:

Choose variables



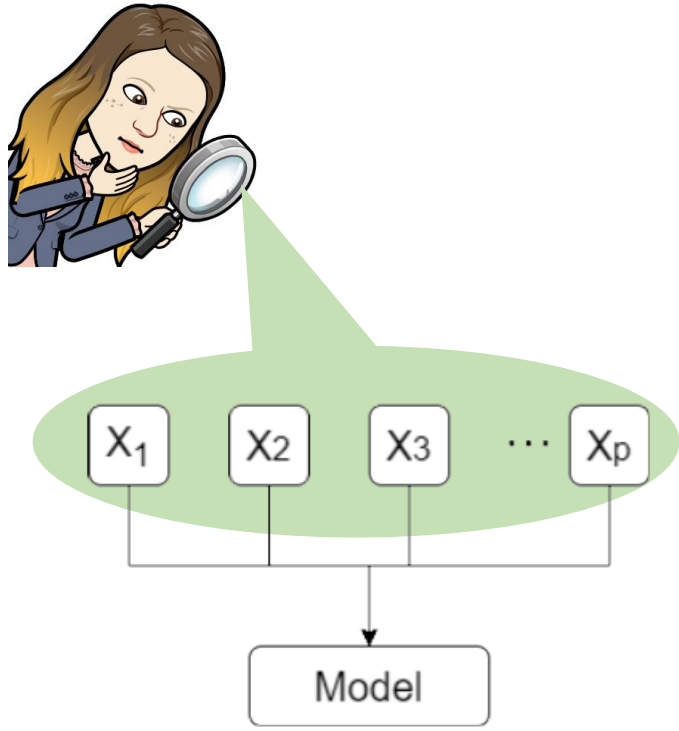
STEP 2:

Check individual covariates using CP diagnostics



STEP 3:

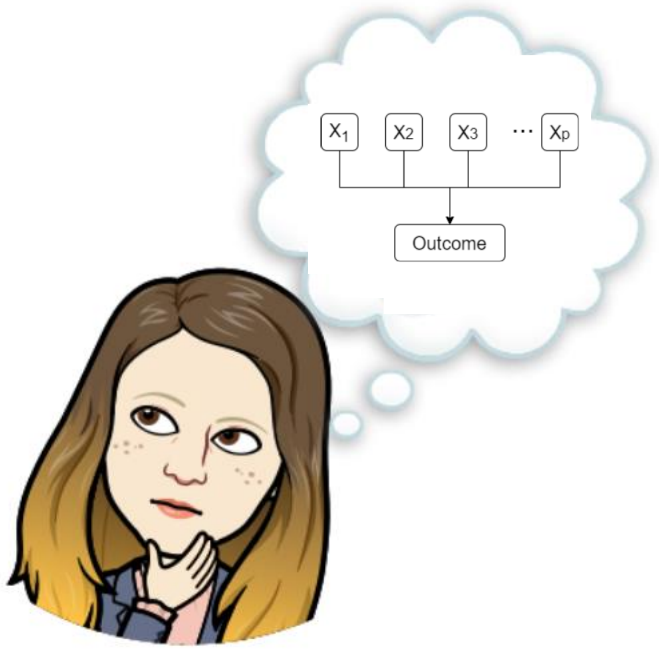
Check overall balance using DRS



So, how best to assess propensity scores?

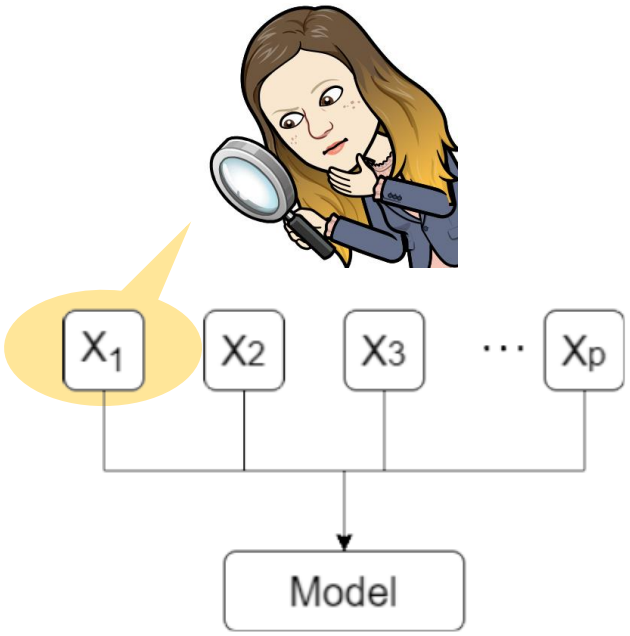
STEP 1:

Choose variables



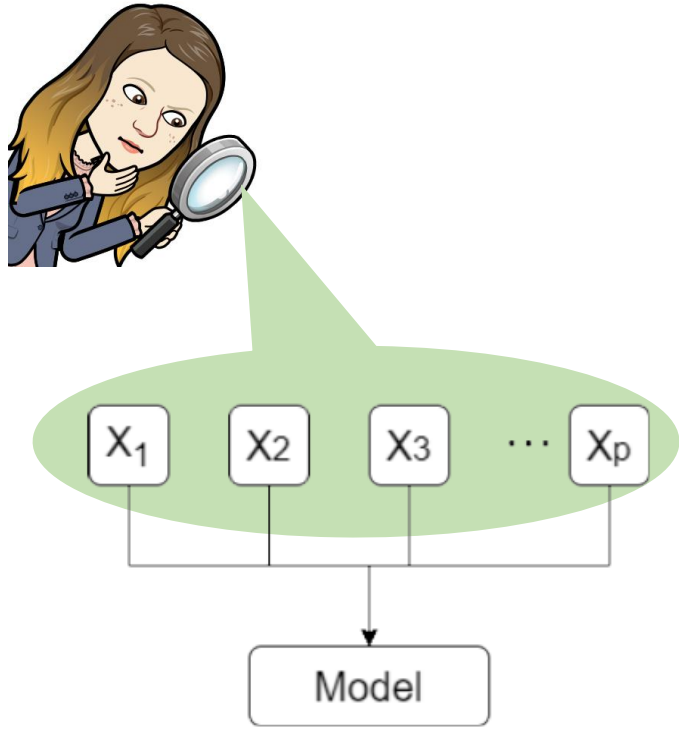
STEP 2:

Check individual covariates using CP diagnostics



STEP 3:

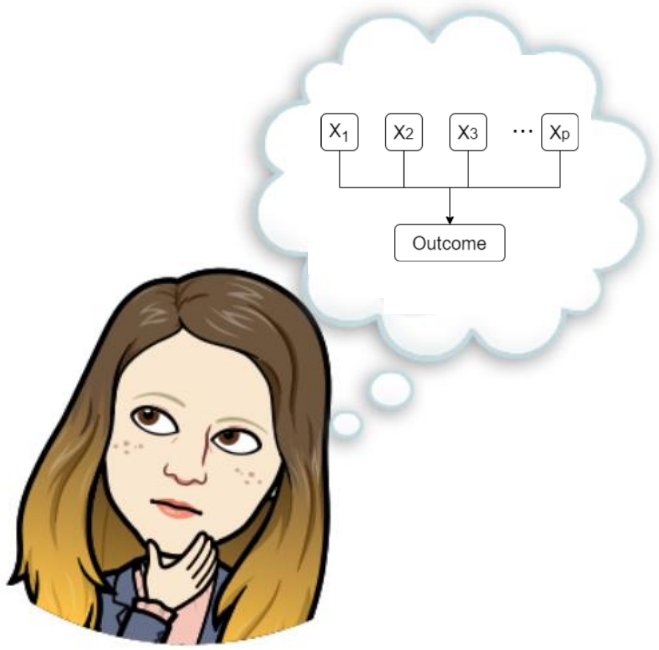
Check overall balance using DRS



So, how best to assess propensity scores?

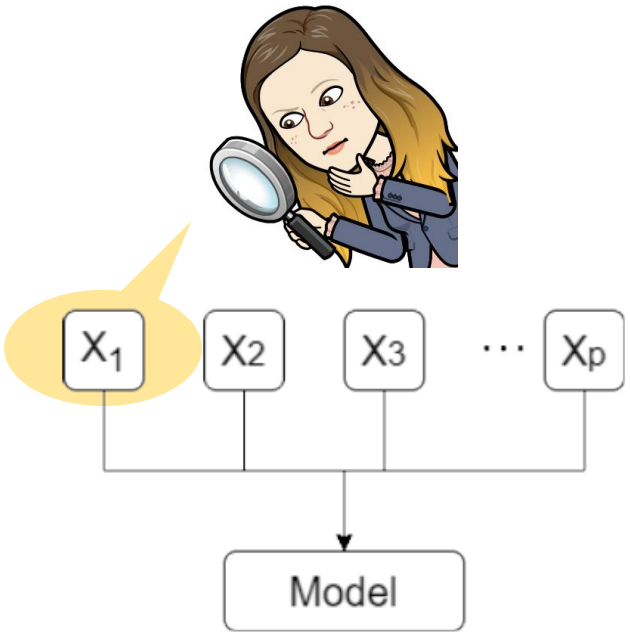
STEP 1:

Choose variables



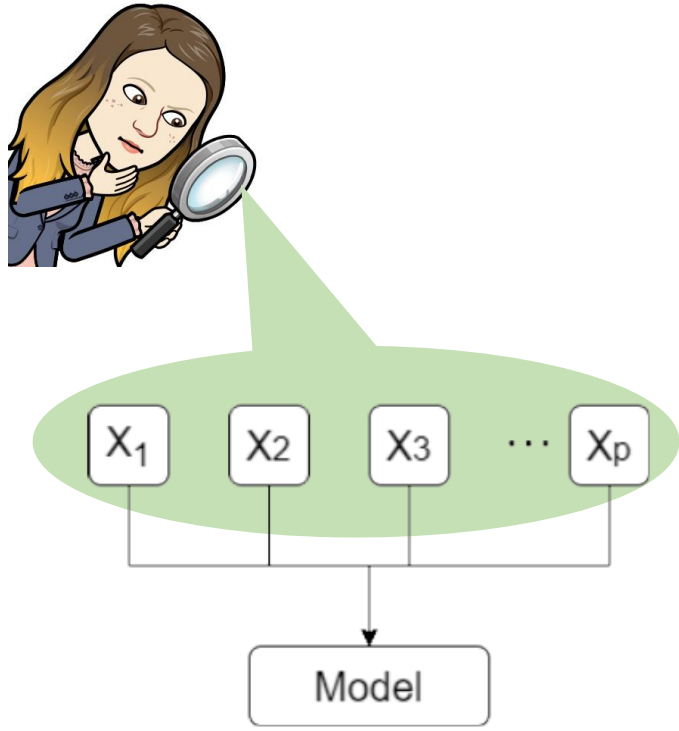
STEP 2:

Check individual covariates using CP diagnostics



STEP 3:

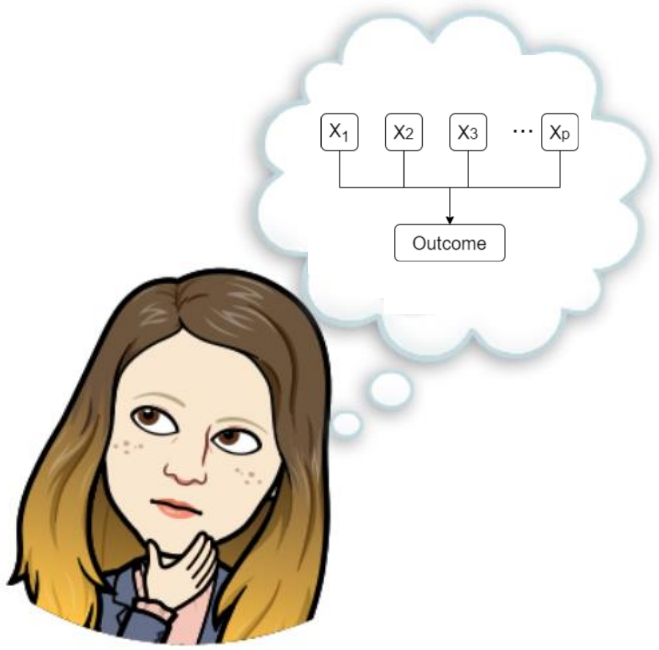
Check overall balance using DRS



So, how best to assess propensity scores?

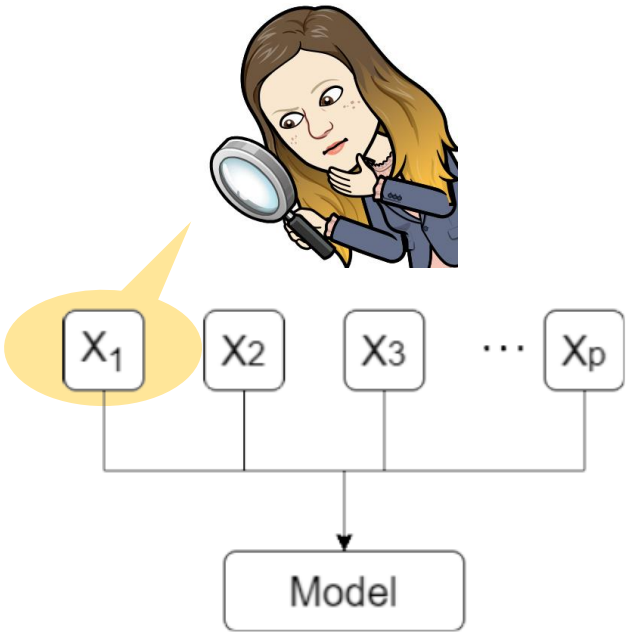
STEP 1:

Choose variables



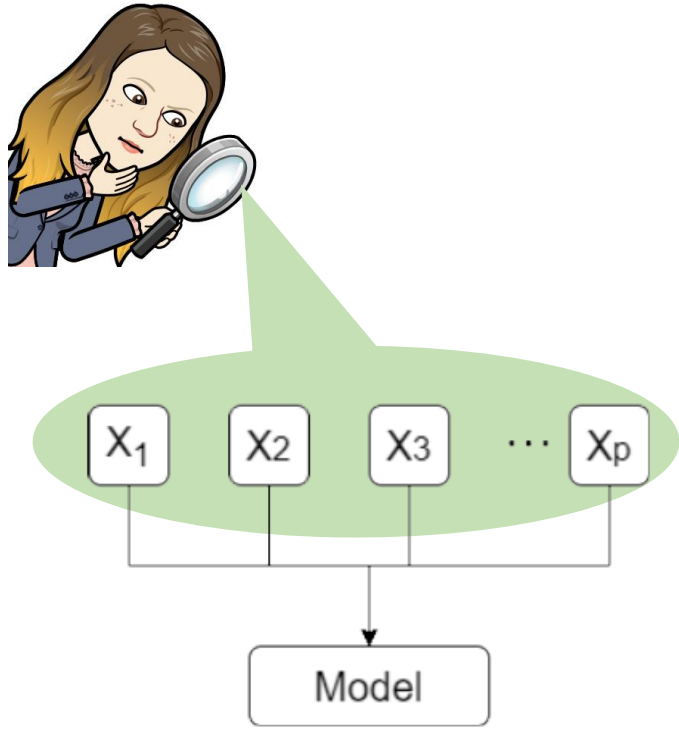
STEP 2:

Check individual covariates using CP diagnostics



STEP 3:

Check overall balance using DRS



Thank you for listening

@EGranger90

MANCHESTER 1824
The University of Manchester

CENTRE FOR EPIDEMIOLOGY VERSUS ARTHRITIS

MRC Medical Research Council

References

[1] Granger, E et al. A review of the use of propensity score diagnostics in papers published in high-ranking medical journals. *BMC Research Methodology.* 2020.

[2] Brookhart, MA et al. Variable selection for propensity score models. *American Journal of Epidemiology.* 2006.

[3] Patrick, AR. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiology and Drug Safety.* 2011.

[4] Caruana, E et al. A new weighted balance measure helped to select the variables to be included in a propensity score model. *Journal of Clinical Epidemiology.* 2015.

[5] Stuart, EA et al. Prognostic score-based balance measures for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology.* 2013

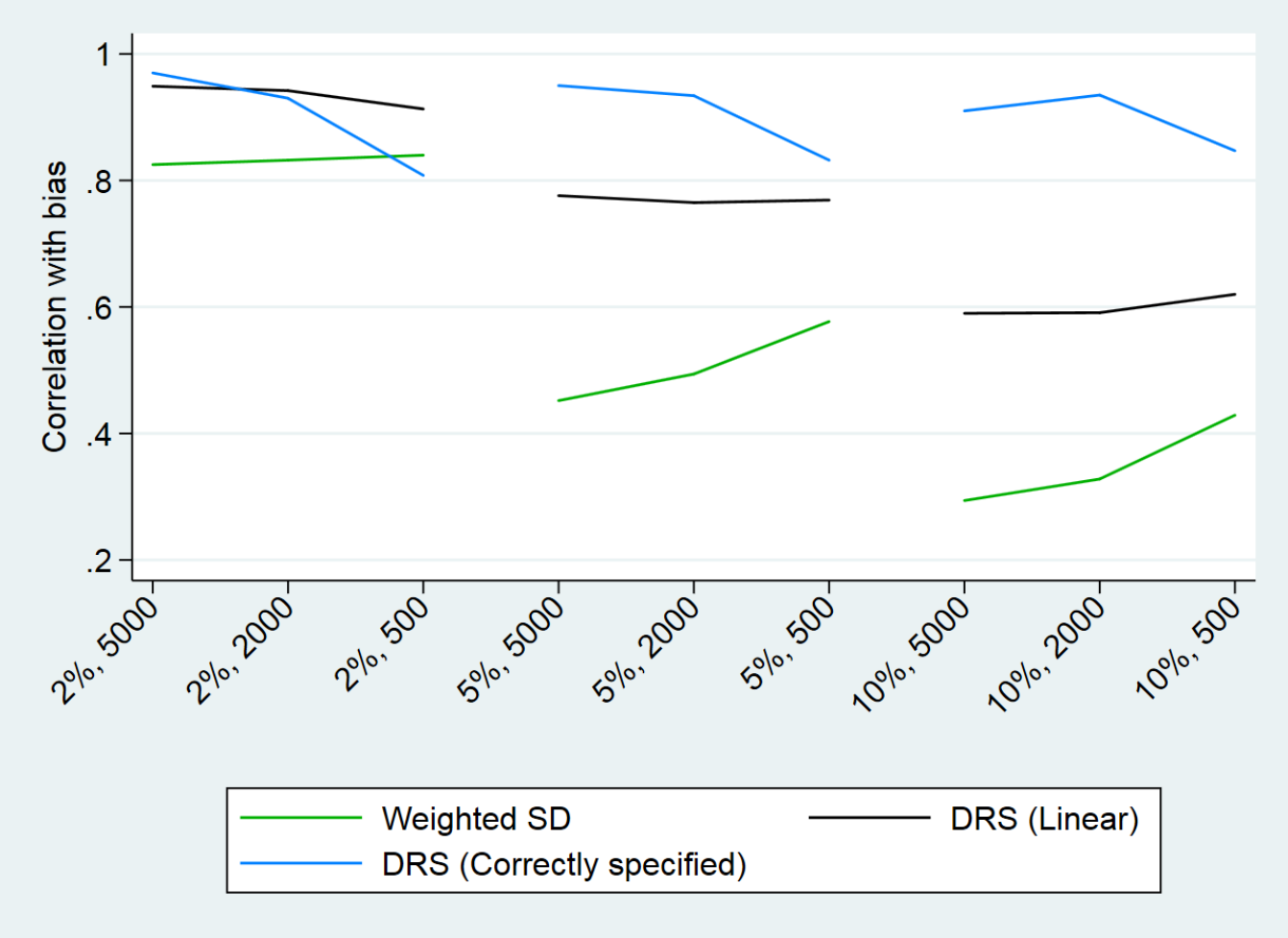
MANCHESTER 1824 The University of Manchester

CENTRE FOR EPIDEMIOLOGY VERSUS ARTHRITIS

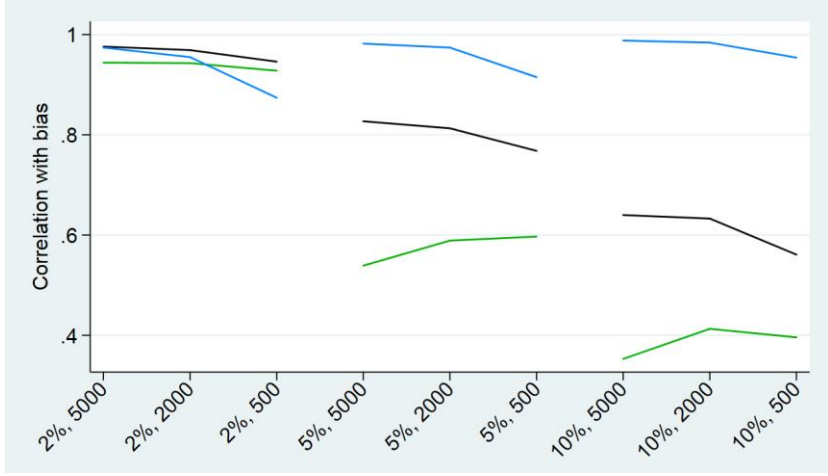
@EGranger90

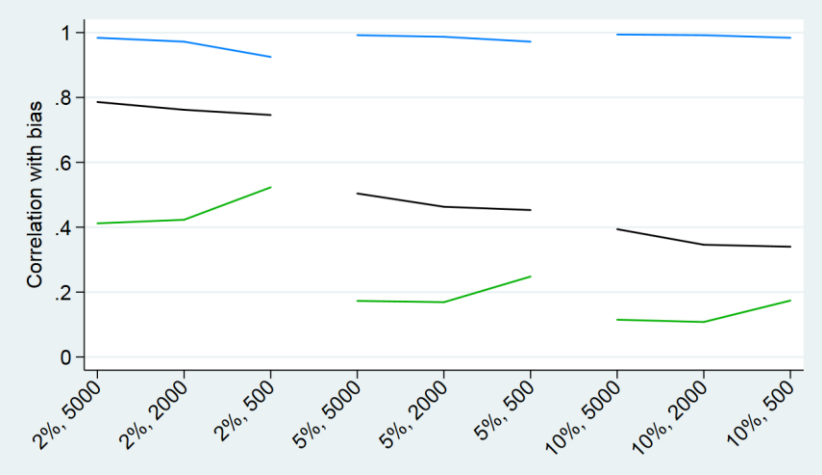
MRC Medical Research Council

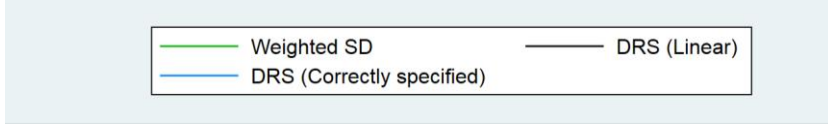
Scenario 2: Non-linear (stratification)



Scenarios 3-5: Interaction terms (stratification)







Figures:

- Scenario 3 (top left) binary-binary
- Scenario 4 (top right) binary-continuous
- Scenario 5 (bottom left) continuous-continuous

Additional weights: Binary outcome

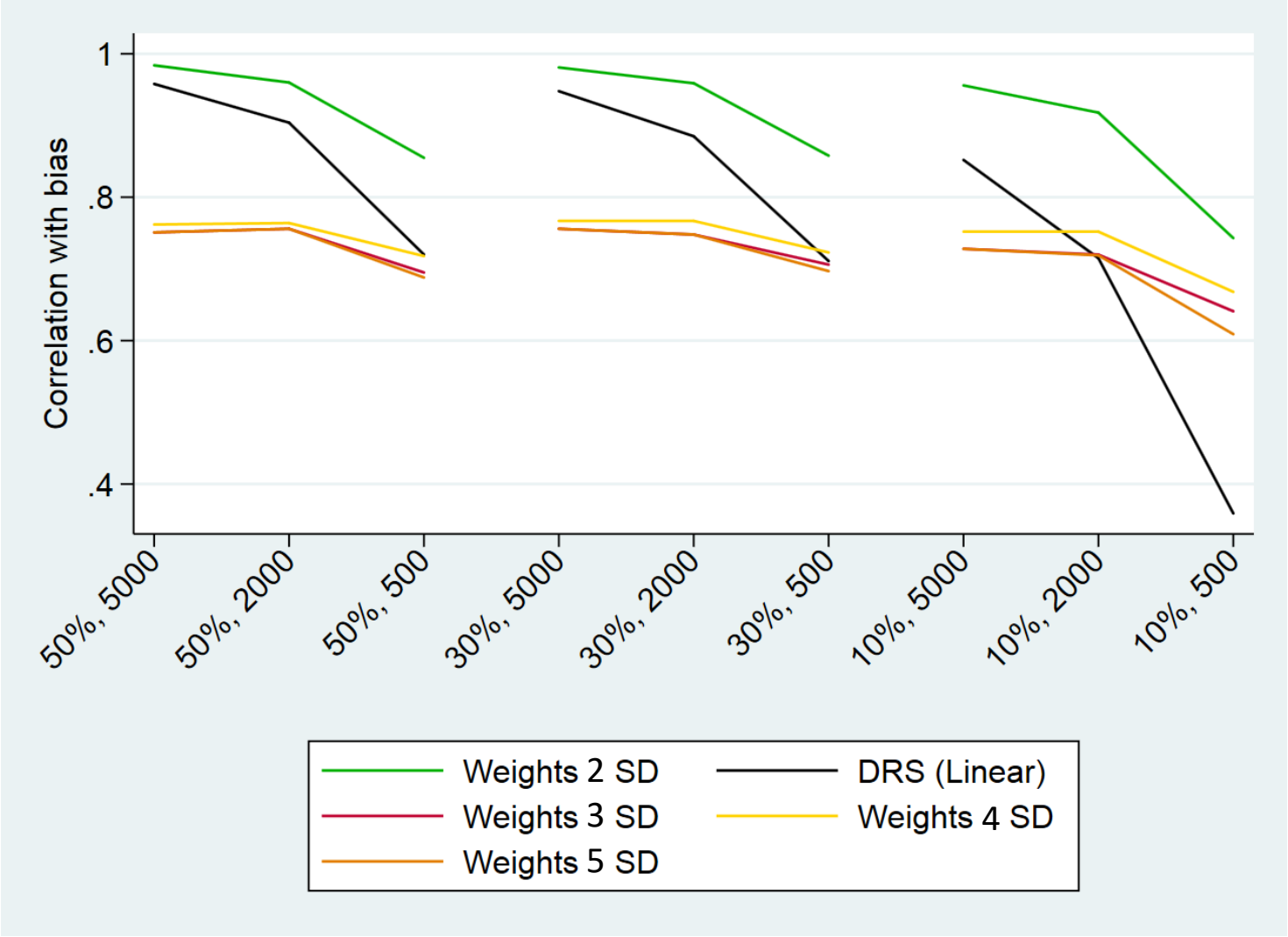
$$w_{3i} = 1 + \log\left(OR_{X_iY}\right) - \frac{1}{p}\sum_{k=1}^{p} \log(OR_{X_kY})$$

$$w_{4i} = 1 + \sqrt{\log\left(OR_{X_iY}\right)} - \frac{1}{p}\sum_{k=1}^{p} \sqrt{\log\left(OR_{X_kY}\right)}$$

$$w_{5i} = 1 + \left|\log\left(OR_{X_iY}\right)\right| - \frac{1}{p}\sum_{k=1}^{p} \left|\log\left(OR_{X_kY}\right)\right|$$

Belitser, SV et al. Measuring balance and model selection in propensity score methods. *Pharmacoepidemiology and Drug Safety.* 2011

Additional scenario: Binary outcome (matching)



Additional scenario: Binary outcome (stratification)