Can We Train Machine Learning Methods to Outperform the High-Dimensional Propensity Score Algorithm?

M. Ehsan. Karim; UBC

2022 Sentinel Innovation and Methods Seminar Series

May 11, 2022

Welcome to the Sentinel Innovation and Methods Seminar Series

The webinar will begin momentarily

Please visit www.sentinelinitiative.org for recordings of past sessions and details on upcoming webinars.

Note: closed-captioning for today's webinar will be available on the recording posted at the link above.

Sentinel

Outline

Slides at tinyurl.com/hdps2022

1. hdPS

- Basic terminology

2. Machine learning-based hdPS

- Karim et al. 2018 Epidemiology
- Joint work with
  - Menglan Pang and Robert W Platt
  - McGill, CNODES Methods; Fund CIHR, Grant #DSE – 146021
- General idea

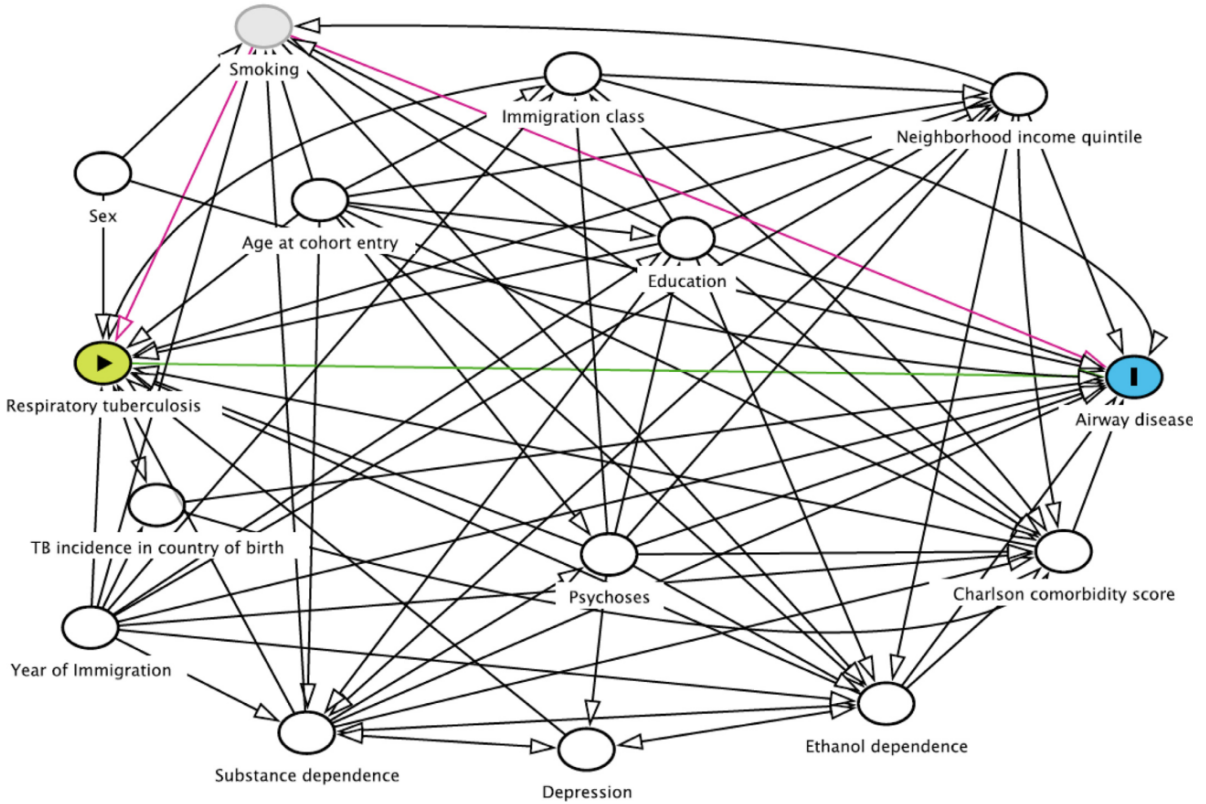
3. Related research

- Not exhaustive

hdPS

Motivating Example

Basham et al. 2021 EClinicalMedicine: CC BY license



Healthcare claims data for immigrants to British Columbia, Canada, 1985–2015

Health care database: Advantages vs Disadvantages

1. Larger sample size;
2. Diverse population;
3. Longitudinal records /many years;
4. Detailed health encounters, comorbidity, drug exposure history;
5. possibility to link other databases.

1. Not specifically designed for answering a particular research question;
2. Data sparsity: relies on visits and encounters;
3. No control over which factors were measured.

TLDR: May not have all confounders.

How to select adjustment variables?

Modified disjunctive cause criterion

Adjust for variables that are

- causes of exposure or outcome or both,
- discard: known instrument,
- including good proxies for unmeasured common causes

VanderWeele et al. 2019 European Journal of Epidemiology: CC BY license

- U = Smoking
- C_1 = Tobacco use

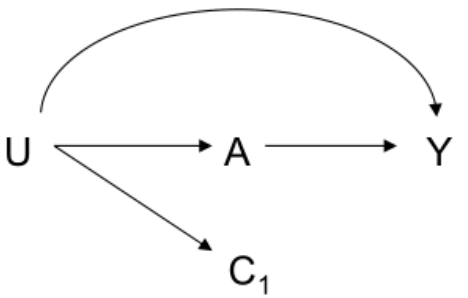
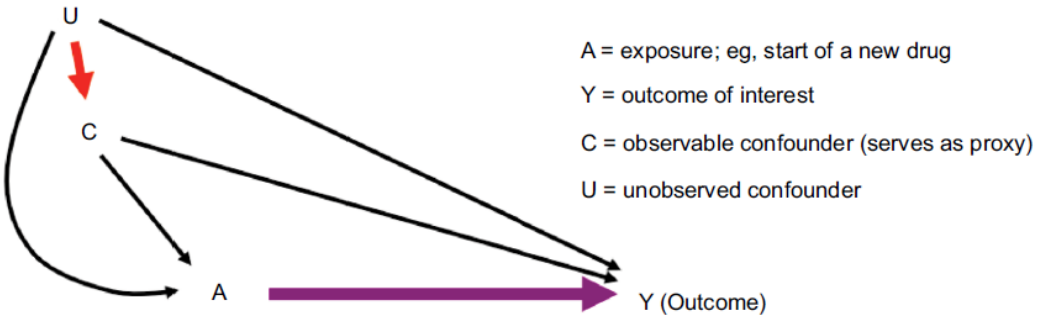


Fig. 5 Control for a proxy confounder C_1 of the true unmeasured confounder U will often, but not always, reduce confounding bias in the relationship between exposure A and outcome Y

Proxy information in Admin data

Schneeweiss et al. 2018 Clinical Epidemiology: CC BY NC license

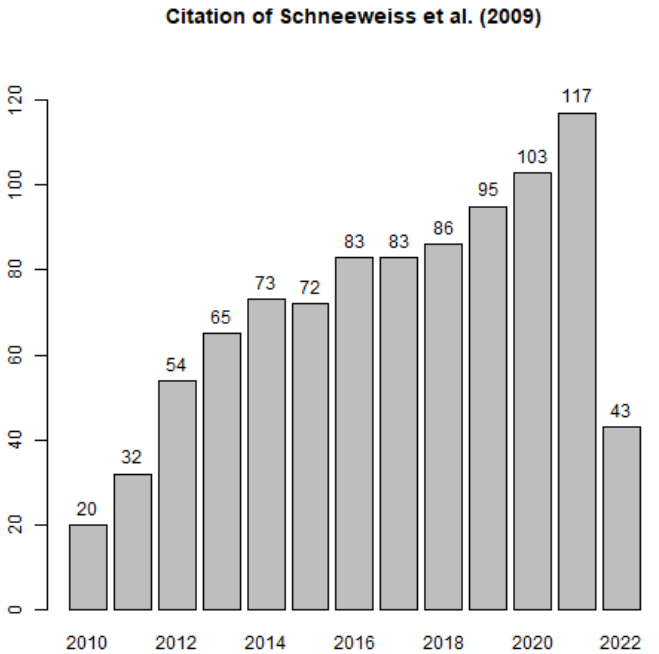
Regular epidemiological studies vs. Proxies of underlying confounders



Unobserved confounder	Observable proxy measurement	Coding examples
Very frail health	Use of oxygen canister	CPT-4
Sick but not critical	Code for hypertension during a hospital stay	ICD-9, ICD-10
Health-seeking behavior	Regular check-up visit; regular screening examinations	ICD-9, CPT-4, #PCP visits
Fairly healthy senior	Receiving the first lipid-lowering medication at age 70 years	NDC, ATC, Read
Chronically sick	Regular visits with specialist, hospitalization; many prescription drugs	#specialist visits, NDC, ATC
Outcome surveillance intensity	General markers for health care utilization intensity	#visits, #different drugs

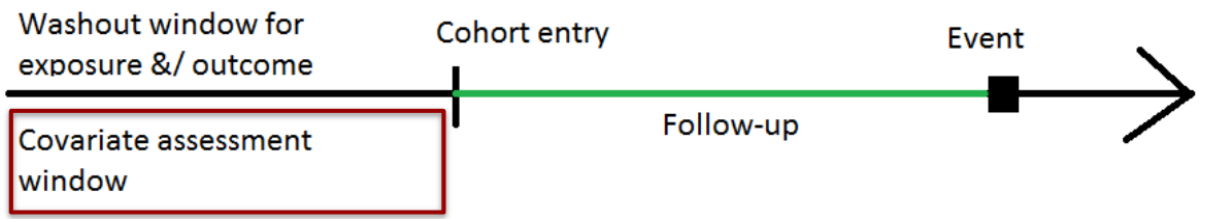
High-dimensional proxy information

- Adjusting for something that **may not be interpretable** directly with the context of the research question.
- Logic: measures from same subject should be **correlated** = relevant proxy information



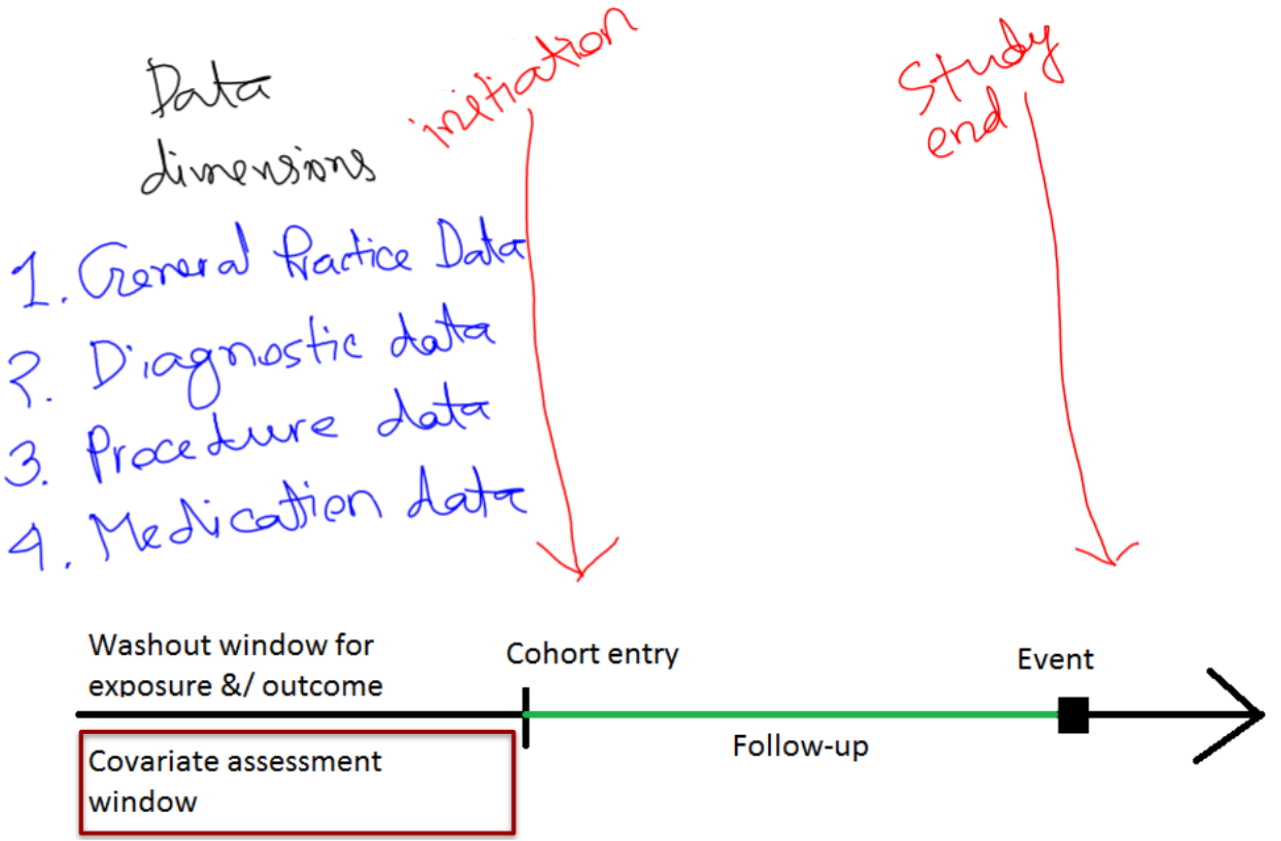
hdPS: General Idea

Karim et al. 2018 Epidemiology: Clinical Practice Research Datalink (1998–2012)



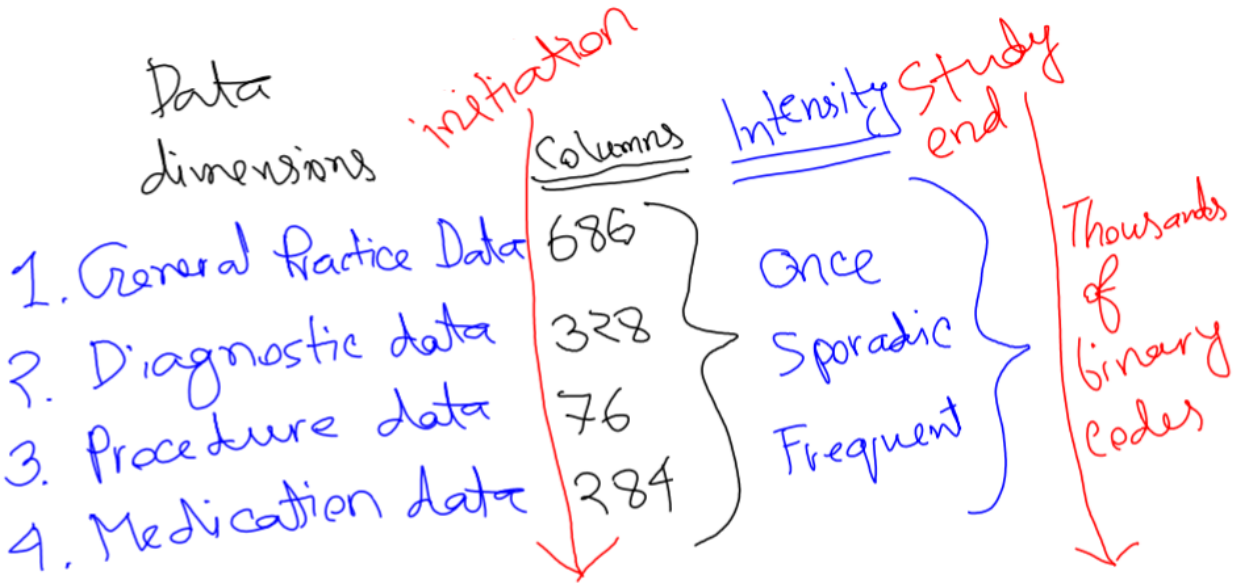
hdPS: General Idea

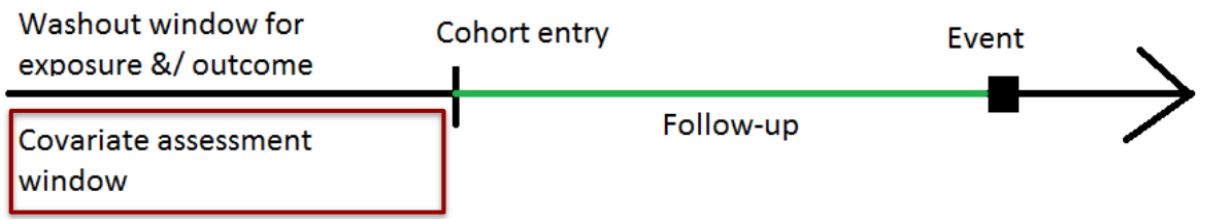
Karim et al. 2018 Epidemiology: Clinical Practice Research Datalink (1998–2012)



hdPS: General Idea

Karim et al. 2018 Epidemiology: Clinical Practice Research Datalink (1998–2012)





hdPS: General Idea

List of additional proxy variables (empirical covariates / EC):

Practice (Dimension 1)	Diagnostic (Dimension 2)	Procedure (Dimension 3)	Medication (Dimension 4)
EC-dim1-1-once	EC-dim2-1-once	EC-dim3-1-once	EC-dim4-1-once
EC-dim1-1-sporadic	EC-dim2-1-sporadic	EC-dim3-1-sporadic	EC-dim4-1-sporadic
EC-dim1-1-frequent	EC-dim2-1-frequent	EC-dim3-1-frequent	EC-dim4-1-frequent
. . .	. . .	. . .	. . .
EC-dim1-686-frequent	EC-dim2-328-frequent	EC-dim3-76-frequent	EC-dim4-284-frequent

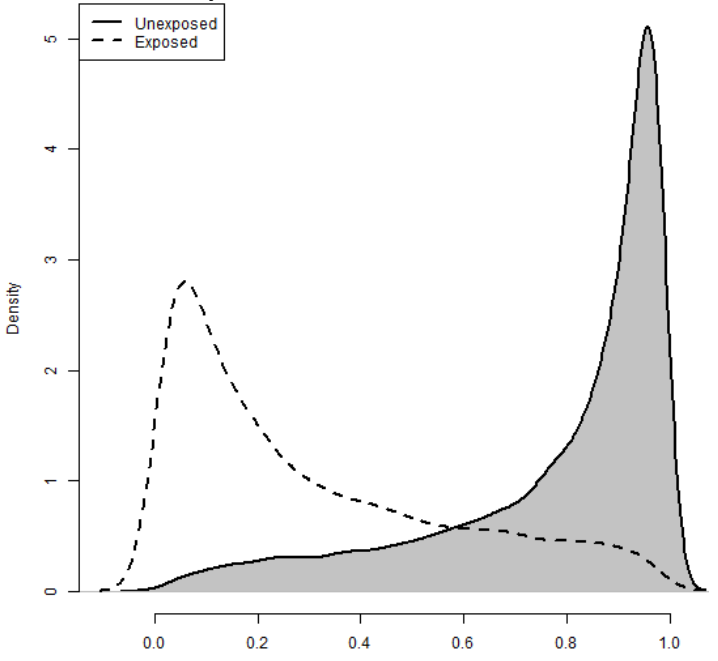
- Total (686+328+76+284)*3 = 4,122 ECs
- 4 dimension $\times$ 3 intensity $\times$ 200 most prevalent codes [*] = **2,400 ECs**
- [*] Schuster et al. (2015) PDS recommended omitting prevalence-based selection

hdPS: General Idea

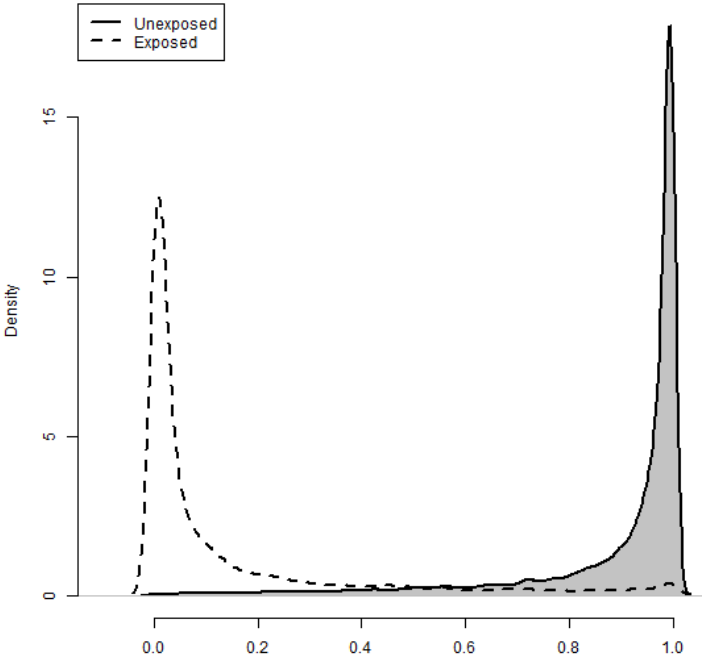
Kitchen Sink Exposure Model (A ~ C + EC)

$$P(A = 1|C, EC) = \frac{1}{1 + exp[\alpha_0 + \alpha_1 C_{\text{important}} + \alpha_2 C_{\text{potential confounder}} + \sum_{i=1}^{2,400} \alpha_i' EC_i]}$$

PS from only baseline confounders



PS from kitchen sink model!



hdPS mechanism: find useful ECs

Assumption:

- $p_{u=1,a=1}$ = prevalence of unmeasured confounder among treated ($A = 1$)
- $p_{u=1,a=0}$ = prevalence of unmeasured confounder among untreated ($A = 0$)
- $p_{u=1,y=1}$ = prevalence of unmeasured confounder among dead ($Y = 1$)
- $p_{u=1,y=0}$ = prevalence of unmeasured confounder among alive ($Y = 0$)

Bross (1966) formula says, the amount of bias due to u is

$$\text{Bias}_M = \frac{p_{u=1,a=1} \times \left(\frac{p_{u=1,y=1}}{p_{u=1,y=0}} - 1\right) + 1}{p_{u=1,a=0} \times \left(\frac{p_{u=1,y=1}}{p_{u=1,y=0}} - 1\right) + 1}$$

In our example,

$$U = \quad \text{smoking status}$$

- Bross (1966) formula requires
  - binary U
  - binary Y
  - binary A

hdPS mechanism: find useful ECs

~~Assumption~~ Calculate:

- $p_{EC=1,a=1}$ = prevalence of ~~unmeasured confounder~~ EC among treated ($A = 1$)
- $p_{EC=1,a=0}$ = prevalence of ~~unmeasured confounder~~ EC among untreated ($A = 0$)
- $p_{EC=1,y=1}$ = prevalence of ~~unmeasured confounder~~ EC among dead ($Y = 1$)
- $p_{EC=1,y=0}$ = prevalence of ~~unmeasured confounder~~ EC among alive ($Y = 0$)

Bross (1966) formula says, the amount of bias due to EC is

$$\text{Bias}_M = \frac{p_{EC=1,a=1} \times \left(\frac{p_{EC=1,y=1}}{p_{EC=1,y=0}} - 1\right) + 1}{p_{EC=1,a=0} \times \left(\frac{p_{EC=1,y=1}}{p_{EC=1,y=0}} - 1\right) + 1}$$

In our example,

$$\begin{aligned} EC = &\quad \text{EC-dim1-21-once} \\ = &\quad \text{EC-dim2-95-once} \\ &\quad \dots \\ = &\quad \text{EC-dim4-64-once} \end{aligned}$$

- Bross (1966) formula requires
  - binary EC
  - binary Y
  - binary A

hdPS mechanism: find useful ECs

Rank (descending) each EC by the magnitude of log-bias: Absolute log $Bias_M$

Rank by bias	Absolute log $Bias_M$	EC
1	0.42	EC-dim1-21-once
2	0.32	EC-dim2-95-once
3	0.25	EC-dim4-289-once
...	...	...
2,400	0.01	EC-dim4-64-frequent

Take top 100 or 500 of these ECs. These are hdPS variables.

hdPS Exposure Model (A ~ C + top-ranked EC)

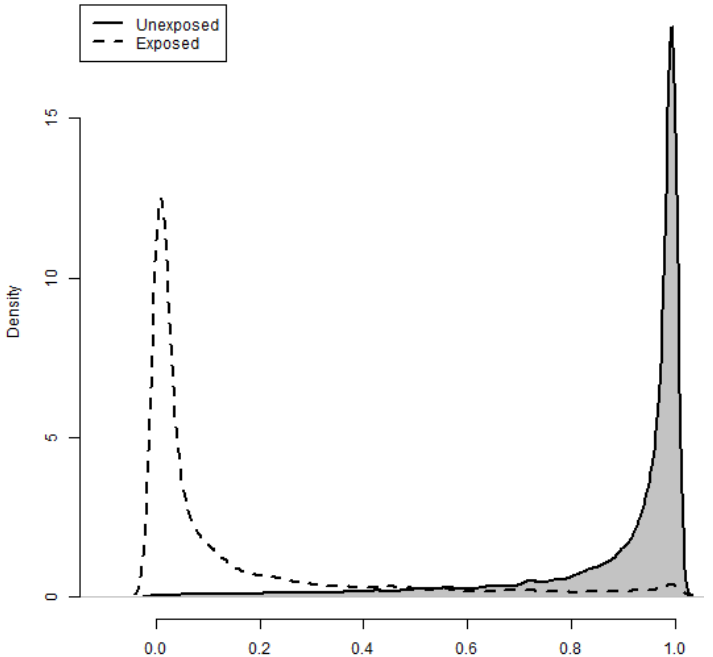
$$P(A=1|C,EC) = \frac{1}{1+exp[\alpha_0 + \alpha_1 C_{\text{important}} + \alpha_2 C_{\text{potential confounder}} + \sum_{i=1}^{\text{top } 500} \alpha_i' EC_i]}$$

hdPS: Assumption

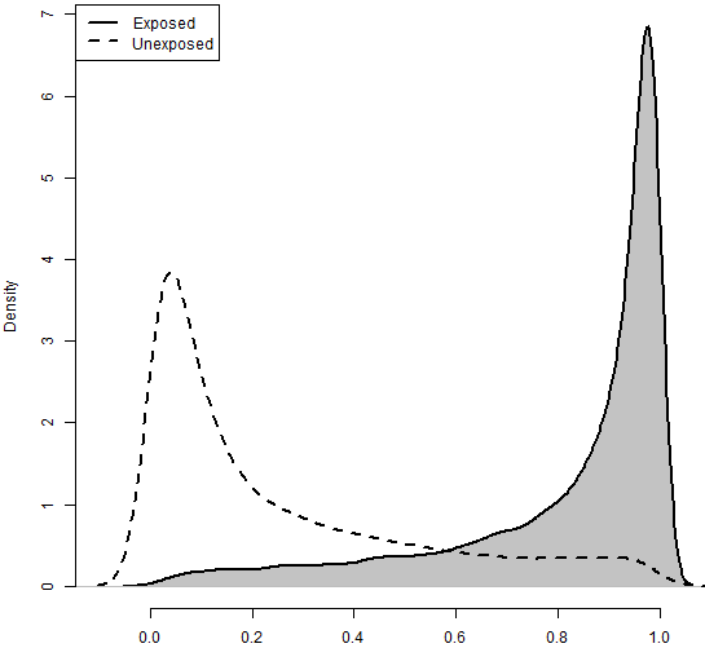
- **The selected ECs collectively serve as proxies of all unmeasured or residual confounding**
- **Implication**: an hdPS analysis may adjust for the unmeasured or residual confounding
- This assumption is strong and often not verifiable.
- Helpful in practice?

hdPS: Balance

PS from kitchen sink model!

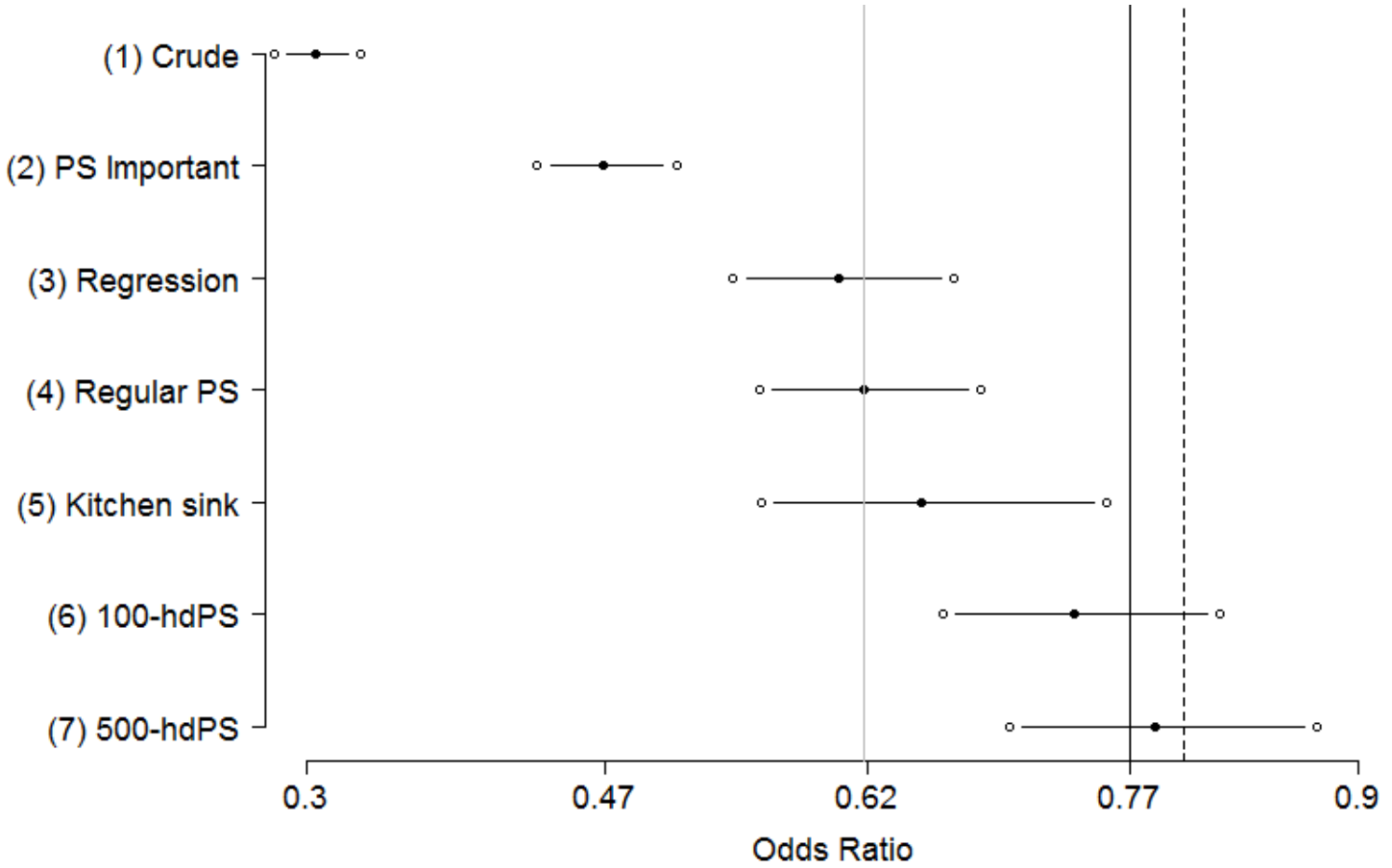


PS from 500-hdPS!



hdPS: estimate treatment effect

- Karim et al. 2018 Epidemiology
- Previous research: Pang et al. (2016): Epidemiology



hdPS: Ways to improve

Rank by bias	Absolute log $Bias_M$	EC
1	0.42	EC-dim1-21-once
2	0.32	EC-dim2-95-once
3	0.25	EC-dim4-289-once
. . .	. . .	. . .
500	0.03	EC-dim4-63-frequent

- ECs selected separately / univatiately VanderWeele et al. 2019 EJE
  - can be **correlated** (coming from same patient),
    - providing same information
    - **may not be useful anymore** in the presence of others
- Multivariate structure is good to consider
  - Model-specification

Machine learning-based hdPS

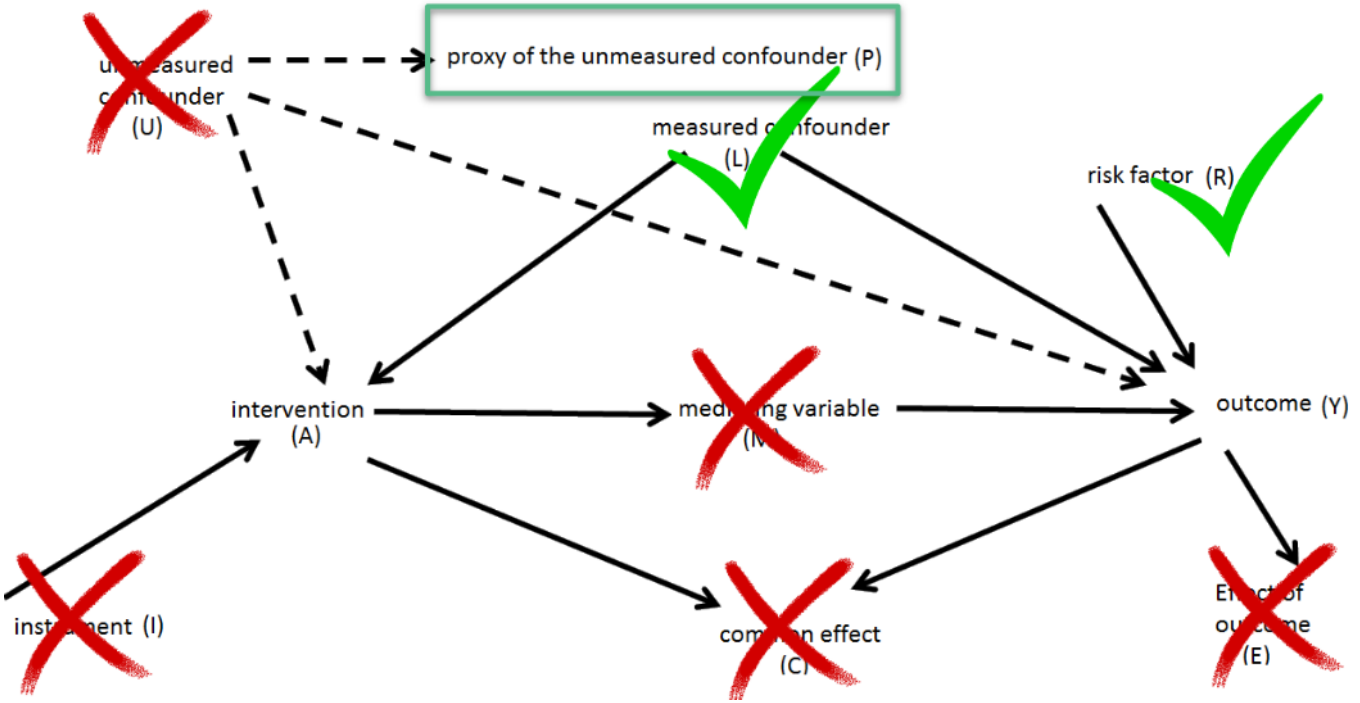
Variable selection in PS context

Literature

- Brookhart et al. (2006) AJE
- Myers et al. (2011) AJE
- Pearl (2011) AJE
- Schuster et al (2016) JCE

- bias amplification
- inflated variance
- overfitting

Variable selection in PS context

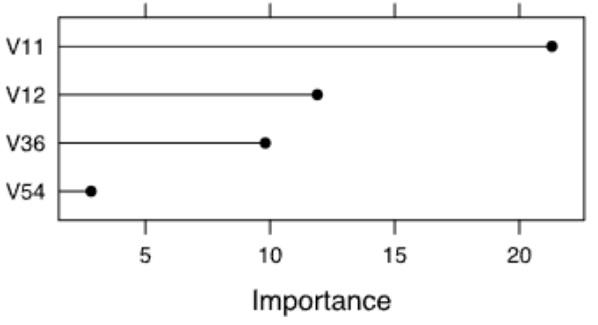


- How to select variables to adjust?
- Same idea for the proxies.
- Pre-exposure measurements (no mediator, collider, effect).
- Associated with Y (irrespective of association with A)

Variable selection via ML

- Jointly consider in 1 model:
  - Perform variable selection based on association with outcome

Approach	Advantage	Limitations
LASSO Franklin et al. (2015) AJE	Variable selection by dropping collinear variables	Tends to select one variable from a group, ignoring the rest
Elastic net	More stable than LASSO	Non-linear and non-additive terms need to be specified
Random forest Low et al. (2016) J. Comp. Eff. Res.	Automatically detect non-linearity and non-additivity	Only provides **variable importance**, but no cut-points



Machine learning-based hdPS

Pure ML approach

Start with all ECs

Outcome Model for EC selection (Y ~ C + ECs)

$$f(Y|C, EC) = \alpha_0 + \alpha_1 C_{\text{important}} + \alpha_2 C_{\text{potential confounder}} + \sum_{i=1}^{2,400} \alpha_i' EC_i$$

Say, 100 ECs (associated with Y) were selected by Elastic net approach

Refined Exposure Model (A ~ C + selected EC)

$$P(A = 1|C, EC) = \frac{1}{1 + exp[\alpha_0 + \alpha_1 C_{\text{important}} + \alpha_2 C_{\text{potential confounder}} + \sum_{i=1}^{\text{selected } 100} \alpha_i' EC_i]}$$

Machine learning-based hdPS

Hybrid approach (hdPS, then ML)

Start with top 500 ECs selected by Bross formula / prioritization

Outcome Model for EC selection (Y ~ C + top-500 ECs)

$$f(Y|C, EC) = \alpha_0 + \alpha_1 C_{\text{important}} + \alpha_2 C_{\text{potential confounder}} + \sum_{i=1}^{500} \alpha_i' EC_i$$

Say, 100 ECs (associated with Y) were selected by Elastic net approach

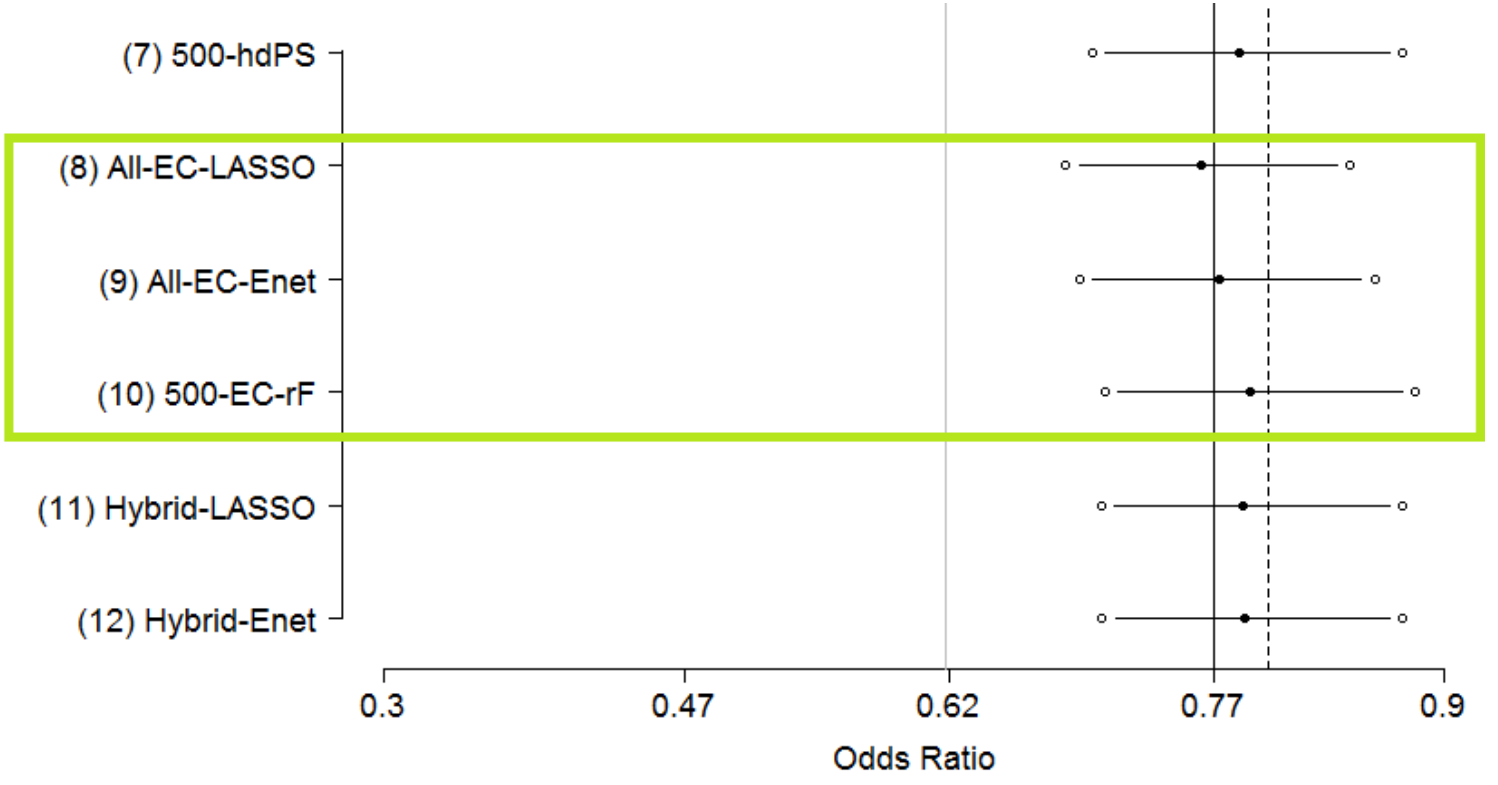
Refined Exposure Model (A ~ C + selected EC)

$$P(A = 1|C, EC) = \frac{1}{1 + exp[\alpha_0 + \alpha_1 C_{\text{important}} + \alpha_2 C_{\text{potential confounder}} + \sum_{i=1}^{\text{selected } 100} \alpha_i' EC_i]}$$

This approach is different than Schneeweiss et al. (2017) Epidemiology, where prioritization was used after applying LASSO.

hdPS vs. ML: estimate treatment effect

Karim et al. 2018 Epidemiology

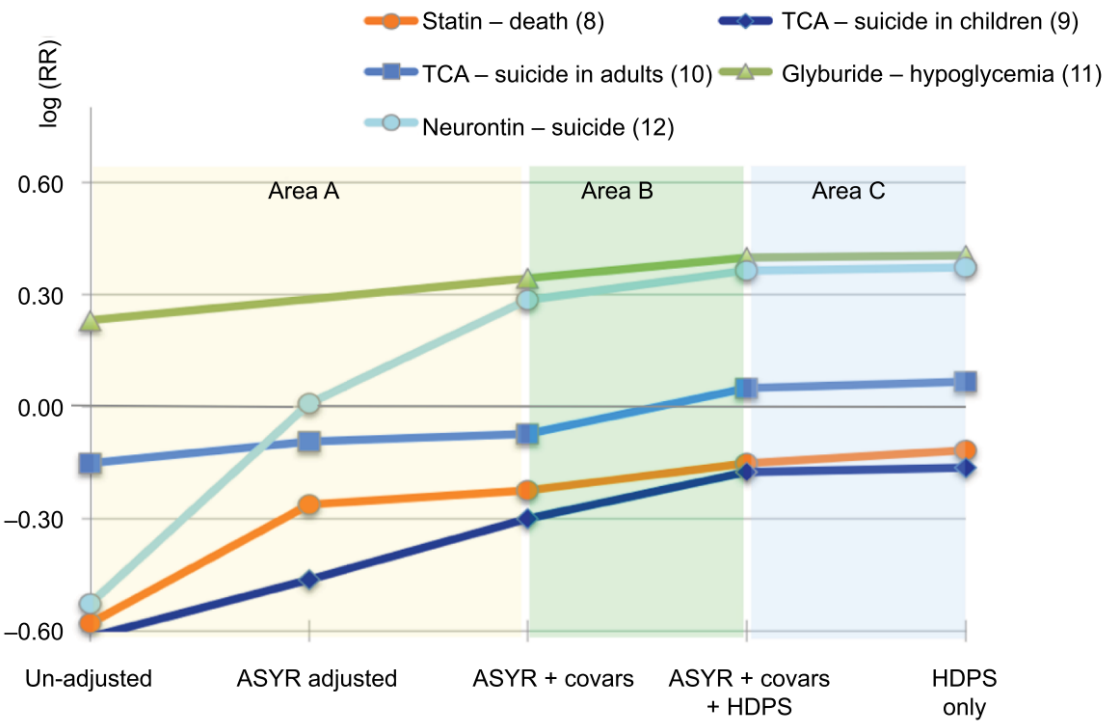


Only ~ 30% of the selected ECs were common.

hdPS: estimate treatment effect

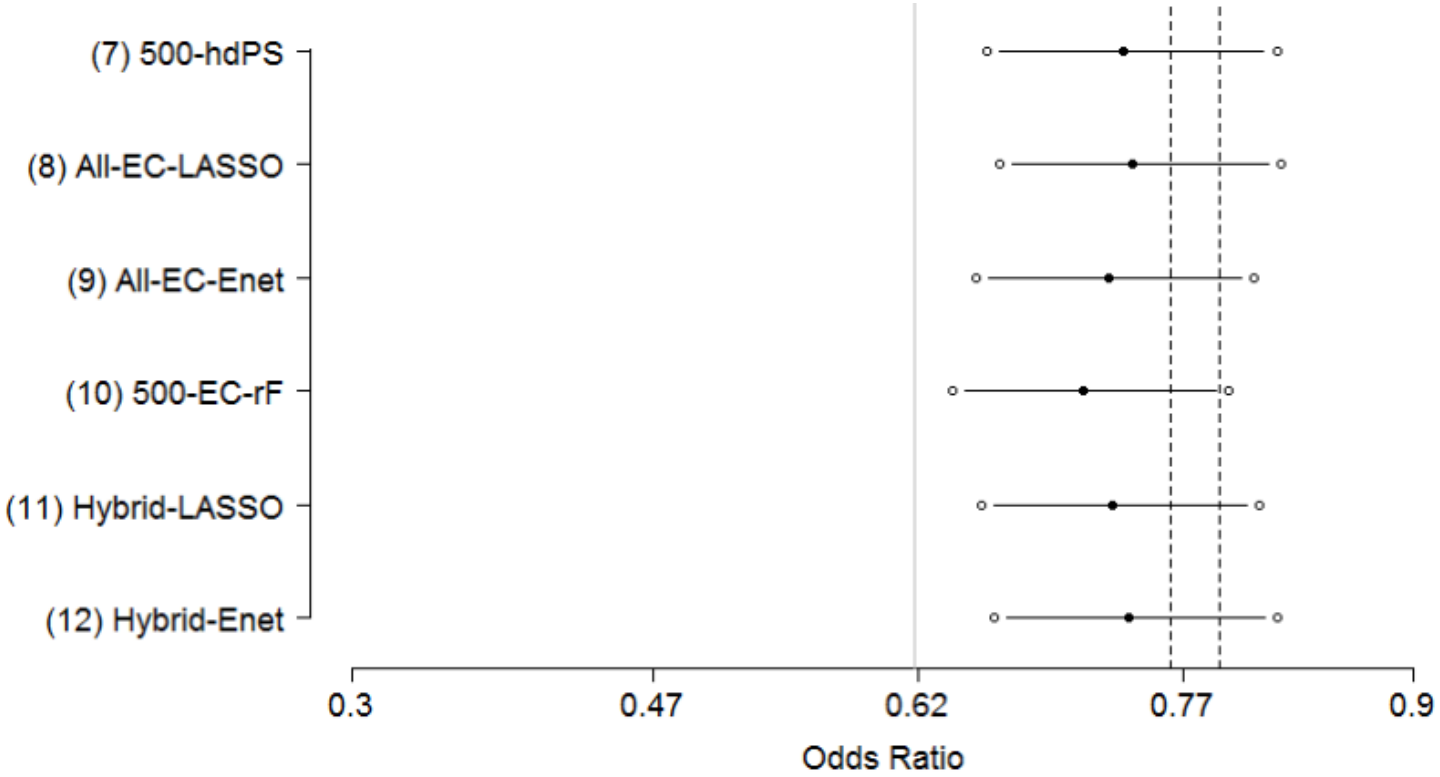
Schneeweiss et al. 2018 Clinical Epidemiology: CC BY NC license

"This strongly suggests that even without the investigator-specifying covariates for adjustment, the algorithm alone optimizes confounding adjustment."



hdPS vs. ML: estimate treatment effect

Karim et al. 2018 Epidemiology



Quality of proxy information matters.

Plasmode Simulation

Franklin et al. (2014) CSDA

scenario	Multiplier of confounder effect	Exposure prevalence	Outcome prevalence	Unmeasured confounder
1-U	1	40	5	Yes *
2-U	3	40	5	Yes *
3-U	5	40	5	Yes *
4-U	1	40	10	Yes *
5-U	3	40	10	Yes *
6-U	5	40	10	Yes *
7-U	1	10	5	Yes *
8-U	3	10	5	Yes *
9-U	5	10	5	Yes *

Another baseline set with no unmeasured confounding (1-A to 9-A).

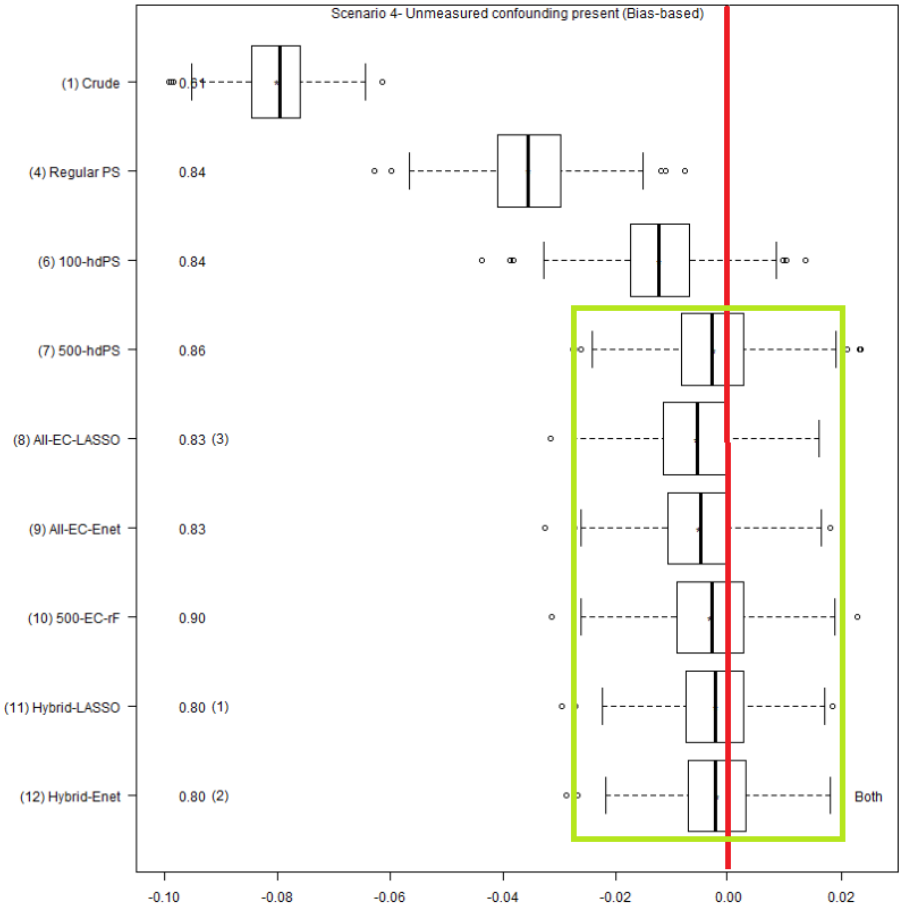
Plasmode Simulation: Leaderboard

Answer to the question in the title of this talk (**bold** = pure ML):

	Bias-Based		Exposure-Based	
Scenario	**MSE**	**Bias**	**MSE**	**Bias**
1-U	Hybrid-Enet	Hybrid-Enet	**All-EC-Enet**	**All-EC-Enet**
2-U	Hybrid-LASSO	*500-hdPS*	**All-EC-Enet**	**All-EC-Enet**
3-U	Hybrid-LASSO	*500-hdPS*	**All-EC-Enet**	**All-EC-Enet**
4-U	Hybrid-Enet	Hybrid-Enet	**500-EC-rF**	**500-EC-rF**
5-U	**500-EC-rF**	**500-EC-rF**	**500-EC-rF**	**500-EC-rF**
6-U	**500-EC-rF**	**500-EC-rF**	**500-EC-rF**	**500-EC-rF**
7-U	Hybrid-Enet	*500-hdPS*	**All-EC-LASSO**	**All-EC-Enet**
8-U	Hybrid-Enet	**500-EC-rF**	**All-EC-LASSO**	**All-EC-LASSO**
9-U	Hybrid-Enet	*500-hdPS*	**All-EC-LASSO**	**All-EC-Enet**
1-A	Hybrid-LASSO	**All-EC-LASSO**	**All-EC-Enet**	**All-EC-LASSO**
2-A	Hybrid-LASSO	Hybrid-LASSO	**All-EC-Enet**	**All-EC-Enet**
3-A	Hybrid-Enet	Hybrid-LASSO	**All-EC-LASSO**	**All-EC-Enet**
4-A	Hybrid-LASSO	**All-EC-Enet**	**All-EC-Enet**	**All-EC-Enet**
5-A	Hybrid-LASSO	**500-EC-rF**	**500-EC-rF**	**500-EC-rF**
6-A	Hybrid-Enet	**500-EC-rF**	**500-EC-rF**	**500-EC-rF**
7-A	Hybrid-Enet	*500-hdPS*	**All-EC-LASSO**	**All-EC-Enet**
8-A	Hybrid-Enet	**500-EC-rF**	**All-EC-LASSO**	**All-EC-Enet**
9-A	Hybrid-LASSO	Hybrid-Enet	**All-EC-LASSO**	**All-EC-Enet**

Plasmode Simulation

Comparable if adequate proxies incorporated (RD estimates)



Shared Limitations

- M-bias Liu et al (2012) AJE
- Z-bias Myers et al. (2011) AJE

- EC interpretation unclear vs. causal inference
  - not collected for research purposes
  - EC used in PS
- Primarily to deal with residual confounding
  - Not a straightforward extension to PS analysis
  - Motivation of PS and hdPS are different to begin with
- No separation of design and analysis stages in bias-based
  - exposure-based is OK; but has own issues
- post-selection bias Taylor and Tibshirani (2015)

Advantage and Limitations

- Alternative ways to prioritize / rank
  - Automatic cut-off of how many variables
  - Ranking
- Pure ML methods can be used for non-binary outcomes and proxies
  - binary
  - categorized
  - continuous
  - survival

- Coverage not assessed Morris et al. (2019)
- Only a few ML methods assessed
- DR methods not covered

Motivating Example

Basham et al. 2021 EClinicalMedicine: CC BY license

Statistical Analysis[a]	N	Adjusted HR	95% CI
Aim 1: analyzing post-TB airway disease risk			
Covariate-adjusted (main analysis: respiratory TB vs controls)	1 005 328	2·08	1·91 – 2·28
Sensitivity analyses			
Covariate-adjusted (removed ETOH, substance dependence, psychoses, and depression)[b]	1 005 328	2·11	1·93 – 2·30
Covariate-adjusted (van Walraven-weighted Elixhauser comorbidity score)[c]	1 005 328	2·06	1·89 – 2·26
Covariate-adjusted (bronchiectasis and fibrosis added to the airway disease definition)	1 005 283	2·18	2·00 – 2·38
Covariate-adjusted (removed respiratory TB patients with pleural samples; $n = 55$)	1 005 273	2·10	1·92 – 2·30
Different TB definitions			
Covariate-adjusted (all forms of TB vs controls)	1 006 271	1·75	1·63 – 1·88
Covariate-adjusted (non-respiratory TB vs controls)[d]	1 004 733	1·36	1·20 – 1·53
Age/sex-adjusted (pleural TB vs non-pleural TB)	1141	0·87	0·57 – 1·32
Aim 2: assessing potential unmeasured confounding			
PS methods			
PS decile-adjusted (main covariates)	1 005 328	2·27	2·08 – 2·49
hdPS decile-adjusted (main covariates + empirical covariates)	1 005 328	2·28	2·09 – 2·50
LASSO-hdPS decile-adjusted (main covariates + LASSO-refined empirical covariates)	1 005 328	2·26	2·07 – 2·47
Adjustment for smoking behavior proxy variables			
Covariate-adjusted subdata analysis (main covariates + tobacco use variable)[e]	31 063	1·53	1·37 – 1·71
Covariate-adjusted (main covariates + personal health risk proxy variable)	1 005 328	2·03	1·85 – 2·22

- Prefer to use hdPS / ML with ECs as a secondary analysis
- Proxy adjustment method (methods vs. subject area journals).

JAMA Example

Brown et al. (2017)

Method	HR	CI 95%
Unadjusted	2.16	1.64-2.86
Regression	1.59	1.17-2.17
IPTW hdPS	1.61	0.997-2.59
1-1 hdPS matching	1.64	1.07-2.53
Pre-pregnancy	1.85	1.37-2.51

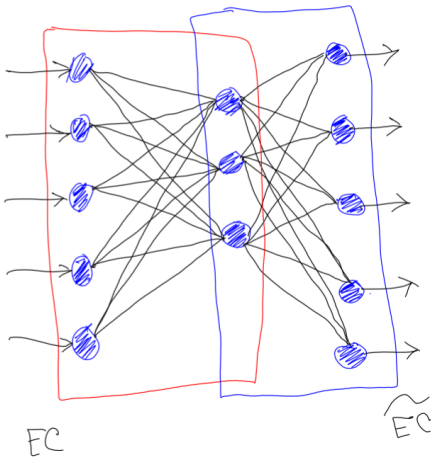
- Conclusion: **not associated**
- More discussion: Amrhein, Trafimow, Greenland, 2019 The American Statistician

Related research directions

Related research directions

AI : Autoencoders

Weberpals et al. (2021) used autoencoders (3, 5, 7 layers) to reduce EC dimensions.



- Autoencoder-based hdPS is useful.
- Shallow learning (less layers) had better MSE.
- Did not perform better than LASSO.

Related research directions

TMLE

Targetted learning approach Pang et al. (2016): Epidemiology

Model	Max SW weight
Only important 5 confounders	1.78
29 confounders	69.67
29 confounders + 400 ECs	390.77

- better covariate balance vs. overfitting
  - Varying number of covariates selected Tazare et al. 2022

Haris and Platt (2021) arxiv

- group importance score
- extension of the hdPS (hdCS) to non-binary outcome and confounders

Related research directions

Sample splitting

Naimi et al. (2021) AJE

SL, TMLE, AIPW and usefulness of sample splitting

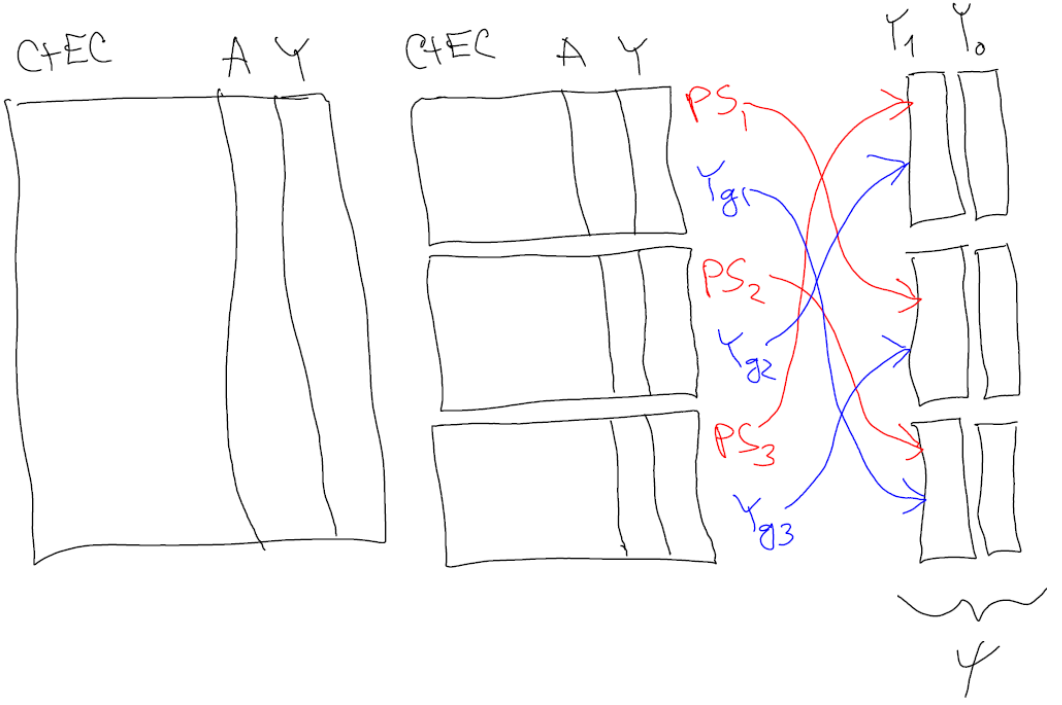
- ML based singly robust methods should be avoided
- Use sample splitting
- rich SL library of flexible regression as well as higher order interactions

Related research directions

Cross-fitting

Zivich and Breskin (2021) Epidemiology

- Cross-fitting + together with double-robust approaches



Related research directions

SL library

Balzer and Westling (2021) AJE

- TMLE without sample-splitting with a carefully chosen SL library

Meng and Huang (2021) arxiv

- SL with smooth (differentiable: LASSO, spline) learners outperform those that included non-smooth learners

Take home message

- hdPS and ML alternatives generally reduces residual confounding
  - [*] if good proxies available.
- hdPS: dependent on Bross-formula (all binary)
- Non-binary outcome: consider ML methods.
- Hybrid-methods performed better (MSE).
- Active area of research

Thanks!

http://ehsank.com/