BIFOLD Welcome Days 19-21th Oct.

Talk by Leila Arras

XAI in Epidemiology

Exploring Interacting Causes of a Health Outcome with CoOL (**C**auses **o**f **O**utcome **L**earning)

BIFOLD

Fraunhofer HHI

Causes of Outcome Learning - paper:

Collaboration with Andreas Rieckmann (Section of Epidemiology,

Dept. of Public Health, University of Copenhagen) & other Co-authors !

Volume 51, Issue 5
October 2022

Article Contents

Abstract

Introduction

The CoOL approach

Real-life application

Discussion

Conclusion

JOURNAL ARTICLE

Causes of Outcome Learning: a causal inference–inspired machine learning approach to disentangling common combinations of potential causes of a health outcome

Andreas Rieckmann ✉, Piotr Dworzynski, Leila Arras, Sebastian Lapuschkin, Wojciech Samek, Onyebuchi Aniweta Arah, Naja Hulvej Rod, Claus Thorn Ekstrøm

International Journal of Epidemiology, Volume 51, Issue 5, October 2022, Pages 1622–1636, https://doi.org/10.1093/ije/dyac078

Published: 08 May 2022 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

Nearly all diseases are caused by different combinations of exposures. Yet, most epidemiological studies focus on estimating the effect of a single exposure on a health outcome. We present the Causes of Outcome Learning approach (CoOL),

XAI Research Group

Dept. of Artificial Intelligence

Fraunhofer HHI

BIFOLD

Fraunhofer HHI

What is Epidemiology?

Definition:

Study of the **distribution** (frequency, pattern) and **determinants** (causes, risk factors) of **health-related states and events** (not just diseases) in **specified populations** (country, global).

Derived from Greek:

epi 'upon, among' + *demos* 'people, district' + *logos* 'study, word, discourse'

= 'the study of what is upon the people'

Not only about epidemic/infectious diseases! But various studies on health-related issues (e.g. pollution, cancer, natural disaster, clinical trials,...).

[www.cdc.gov/careerpaths/k12teacherroadmap/epidemiology.html, en.wikipedia.org/wiki/Epidemiology]

Goal: Discover exposures associated with an outcome

Binary Exposures: X_i

- sex (male/female)
- age category
- taking drug A
- physically active
- smoking
- high BMI
- high blood pressure
- exposed to pollutant A
- ...

Questions:

- **Which exposures** cause an increased risk?
- Do they act **alone or in synergy**?

Health Outcome: Y

Disease (1=Yes or 0=No)

Challenges:
- In practice exposures often interact (no single-cause disease)
- Numerous sources of exposures (individual's genetic characteristics, environmental factors, lifestyle...), potentially all exposures from conception to death (exposome)

[Patel 2017, *Analytic Complexity and Challenges in Identifying Mixtures of Exposures Associated with Phenotypes in the Exposome Era*, Curr Epidemiol Rep, doi.org/10.1007/s40471-017-0100-5]

Synergy - Interaction on an additive scale

Absolute risks of lung cancer:

	$X_1=0$	$X_1=1$
$P(Y=1\|X_1,X_2)$	No Asbestos	Asbestos
$X_2=0$ Non-Smoker	0.11 %	0.67 %
$X_2=1$ Smoker	0.95 %	4.50 %

IC > 0 positive interaction
IC < 0 negative interaction
IC = 0 no interaction

Interaction coefficient:

baseline risk

Defined as difference of risks w.r.t. baseline risk

$$IC = (R_{A\bar{S}} - R_{\bar{A}\bar{S}}) - [(R_{A\bar{S}} - R_{\bar{A}\bar{S}}) + (R_{\bar{A}S} - R_{\bar{A}\bar{S}})]$$
$$= R_{AS} - R_{A\bar{S}} - R_{\bar{A}S} + R_{\bar{A}\bar{S}}$$
$$= 4.5 - 0.67 - 0.95 + 0.11 = 2.99$$

=> positive interaction between smoking and asbestos

=> identify subgroup for public health intervention

[VanderWeele and Knol 2014, *A Tutorial on Interaction, Epidemiol Methods, doi.org/10.1515/em-2013-0005*]

Synergy - Interaction on an additive scale

Absolute risks of lung cancer:

	$X_1=0$	$X_1=1$
$P(Y=1\|X_1,X_2)$	No Asbestos	Asbestos
$X_2=0$ Non-Smoker	0.11 %	0.67 %
$X_2=1$ Smoker	0.95 %	4.50 %

IC > 0 positive interaction
IC < 0 negative interaction
IC = 0 no interaction

=> identify subgroup for public health intervention

Interaction coefficient:

baseline risk

Defined as difference of risks w.r.t. baseline risk

$$IC = (R_{AS} - R_{\bar{A}\bar{S}}) - [(R_{A\bar{S}} - R_{\bar{A}\bar{S}}) + (R_{\bar{A}S} - R_{\bar{A}\bar{S}})]$$
$$= R_{AS} - R_{A\bar{S}} - R_{\bar{A}S} + R_{\bar{A}\bar{S}}$$
$$= 4.5 - 0.67 - 0.95 + 0.11 = 2.99$$

=> positive interaction between smoking and asbestos

Can we discover interaction with machine learning?

[VanderWeele and Knol 2014, *A Tutorial on Interaction, Epidemiol Methods, doi.org/10.1515/em-2013-0005*]

Standard approach: “linear” regression

Model:

$$P(Y|X_1=x_1, X_2=x_2) = c_0 + c_1 \cdot x_1 + c_2 \cdot x_2 + c_3 \cdot x_1 x_2$$

regression terms = all possibles combinations of input variables

where

$$c_0 = P(Y|X_1=0, X_2=0)$$ baseline risk

$$c_1 = P(Y|X_1=1, X_2=0) - P(Y|X_1=0, X_2=0)$$ risk diff due to X1 alone

$$c_2 = P(Y|X_1=0, X_2=1) - P(Y|X_1=0, X_2=0)$$ risk diff due to X2 alone

$$c_3 = P(Y|X_1=1, X_2=1) - P(Y|X_1=1, X_2=0) - P(Y|X_1=0, X_2=1) + P(Y|X_1=0, X_2=0)$$

risk diff due to additive interaction between X1 and X2

[VanderWeele and Knol 2014, *A Tutorial on Interaction, Epidemiol Methods, doi.org/10.1515/em-2013-0005*]

Linear regression for higher-order interactions?

N binary exposures: 2^N regression terms

11 binary exposures: 2048 regression terms

Drawbacks:

- Model overfitting, study not reproducible
- Require large sample
- Computationally challenging
- Hard to interpret interactions with overlapping sets of variables
- Results can be misleading even if p-value of regression coefficients is low

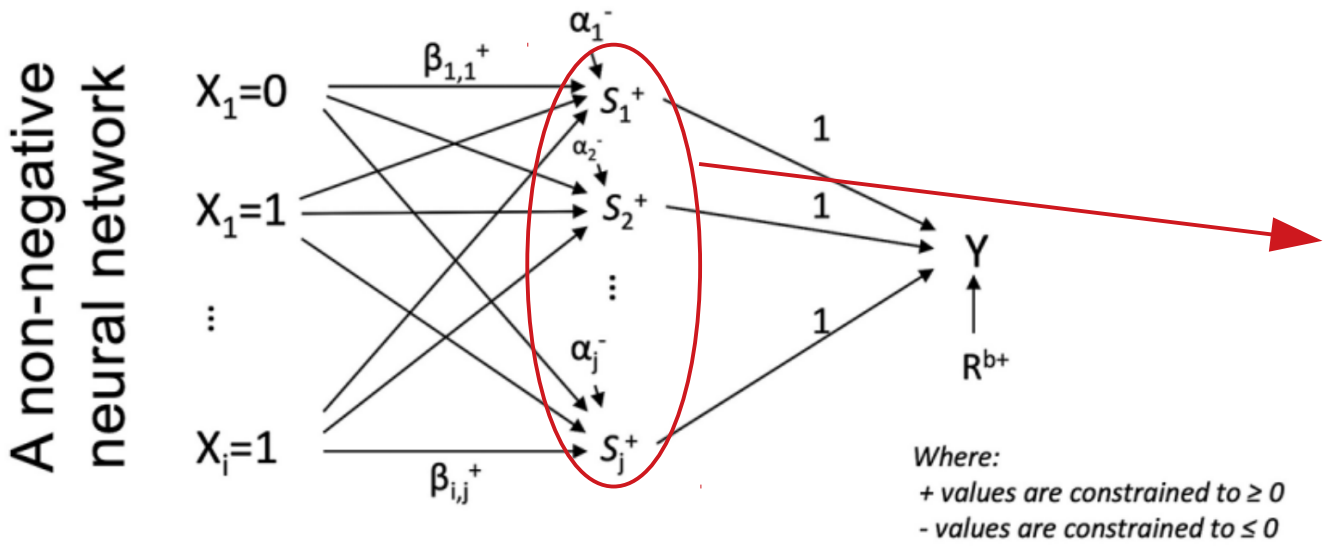
Solution: reduce the number of tested interactions. Alternative: use CoOL

[Patel 2017, *Analytic Complexity and Challenges in Identifying Mixtures of Exposures Associated with Phenotypes in the Exposome Era*, Curr Epidemiol Rep, doi.org/10.1007/s40471-017-0100-5]

Our approach: neural network + XAI + clustering

$$P(Y=1|X_1, X_2, \ldots) = \sum_j \Big(\underbrace{ReLU\ \Big(\sum_i X_i \cdot \beta^+_{i,j} + \alpha^-_j\Big)}_{\text{hidden layer activations } S^+_j} \Big) + R^{b+}$$

baseline risk



Non-linear hidden layer can modelize any higher-order additive interaction (no more need to test all of them explicitly!), as well as standalone exposure effects without interaction

Model assumptions for synergy detection

Positive Monotonicity: **e.g. sufficient-component-cause framework**

Each exposure either increases risk or has no effect for all individuals in the population (i.e. regardless of other exposures) as its value changes from 0 to 1

if "risk diff for X_1 in strata X_2=1" > "risk diff for X_1 in strata X_2=0"
then synergy

Relaxed Monotonicity: **CoOL framework**

Each exposure either increases risk or has no effect on each individual separately (i.e. depending on other exposures) with no pre-defined direction

In practice: one-hot encoding of inputs (even for exposures with 2 categories) allows to discover e.g. that "drug A only harmful for women" and "drug B only harmful for men"

if "combined risk of exposures" > "sum of risks due to standalone exposures"
then synergy

[VanderWeele and Robins 2007, *The Identification of Synergism in the Sufficient-Component-Cause Framework,* Epidemiology, doi.org/10.1097/01.ede.0000260218.66432.88]

Step 1: Model fitting

- Training via Stochastic Gradient Descent (update model one individual at a time)
- Minimize squared prediction error (data loss $\left(Y_{true} - \hat{P}_{Model}\left(Y|\boldsymbol{X}\right)\right)^2$)
- Weight regularization through squared L2-norm penalty, to avoid overfitting on noise (regularization loss $||\boldsymbol{\beta}||^2$)
- Initialization of baseline risk R^{b+} with mean risk of the outcome $E\left[Y_{true}\right]$
- Split data in train & internal validation sets to assess reproducibility of found risk factors
- Even though overall discriminative performance (AUC) is low, the model can still capture important sets of causes for particular subgroups!
  (e.g. improved prediction on subgroups with rare risk factors that have strong effects)

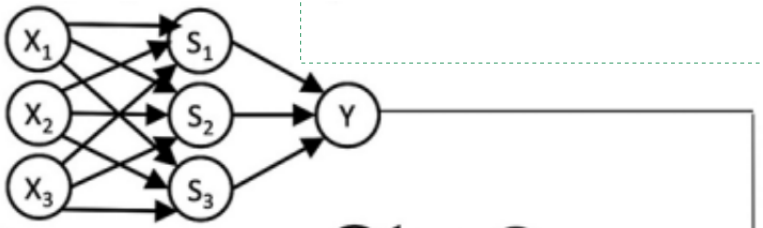
[Janssens and Martens 2020, *Reflection on modern methods: Revisiting the area under the ROC Curve*, Int J Epidemiol, doi.org/10.1093/ije/dyz274]

Step 2: Decompose prediction into risk contributions

For each individual do:

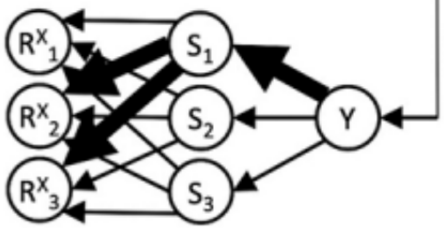
start X_i exposures

1. Forward pass [predict]



$$Y_{predicted} = P_{Model}(Y=1|\boldsymbol{X}) = \sum_j \Big(\underbrace{ReLU\ \big(\sum_i X_i \cdot \beta^+_{i,j} + \alpha^-_j\big)}_{hidden\ layer\ activations\ S^+_j} \Big) + R^{b+}$$

baseline risk



2. Backward pass [explain prediction]

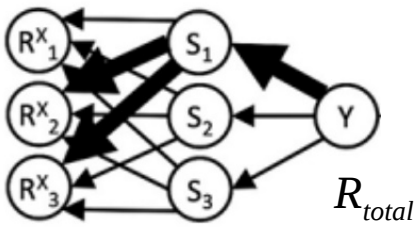
$$P_{Model}(Y=1|\boldsymbol{X}) = R^{b+} + \sum_i R_i$$

end R_i risk contribution for X_i

As an explainable artificial intelligence (XAI) method we use Layer-wise Relevance Propagation (LRP)

[Bach et al. 2015, *On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation*, PLOS ONE, doi.org/10.1371/journal.pone.0130140]

Step 2: Decompose prediction into risk contributions



We use the $LRP_{\alpha=1,\beta=0}$ decomposition rule

Output layer: $$R_{total} = P_{Model}(Y=1|\boldsymbol{X}) - R^{b+}$$

baseline risk

Hidden layer: $$R_j = \frac{S_j}{\sum_{j'} S_{j'}} R_{total}$$

Input layer: $$R_i = \sum_j \frac{X_i \cdot \beta^+_{i,j}}{\sum_{i'} X_{i'} \cdot \beta^+_{i',j}} R_j$$

Advantages:
- overall risk conserved, no risk assigned to the hidden layer intercepts
- explains which exposures might by be “causing” the outcome, rather than what would be the impact of modifying certain exposures (sensitivity-based, perturbation-based XAI)
- model design fully matches the theory behind LRP: deep Taylor decomposition

[Montavon et al. 2017, *Explaining nonlinear classification decisions with deep Taylor decomposition*, Pattern Recognition, doi.org/10.1016/j.patcog.2016.11.008]

Step 3: Clustering risk contributions

- Hierarchical clustering of individuals based on risk contributions (using the Ward's algorithm and Manhattan distances)
- Visualize clustering hierarchy as dendrogram to decide on the number of clusters/subgroups
- For each subgroup compute risk contributions (mean and std)

  1) mean risk vs. prevalence plot
  => "area above baseline" indicates public health impact of subgroup
  2) mean risk table
  => "sum of standalone risks < combined risk" indicates synergism

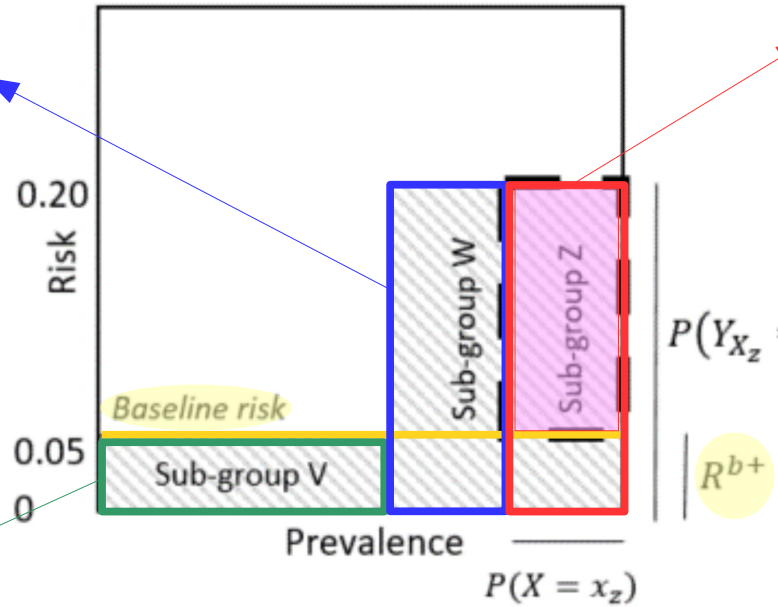
[Strauss and Maltitz 2017, *Generalising Ward's Method for Use with Manhattan Distances*, PLOS ONE doi.org/10.1371/journal.pone.0168288]

Step 3: Clustering risk contributions

Mean risk vs. prevalence by subgroup

another subgroup W with increased risk

a subgroup Z with increased risk
i.e. with mean risk > baseline risk
and a corresponding set of exposures x_z
(typically low prevalence)



$P(Y_{X_z} = 1)$ | $\sum_i R_i^{X_z}$

R^{b+}

baseline subgroup V
with mean risk = baseline risk
(typically largest prevalence)

$P(Y = 1) =$ [ ]

Excess fraction for subgroup Z:

$$\frac{P(Y=1|\boldsymbol{X}=\boldsymbol{x_z}) - R^{b+}}{P(Y=1)} = \frac{\square}{\square}$$

=> public health impact

Step 3: Clustering risk contributions

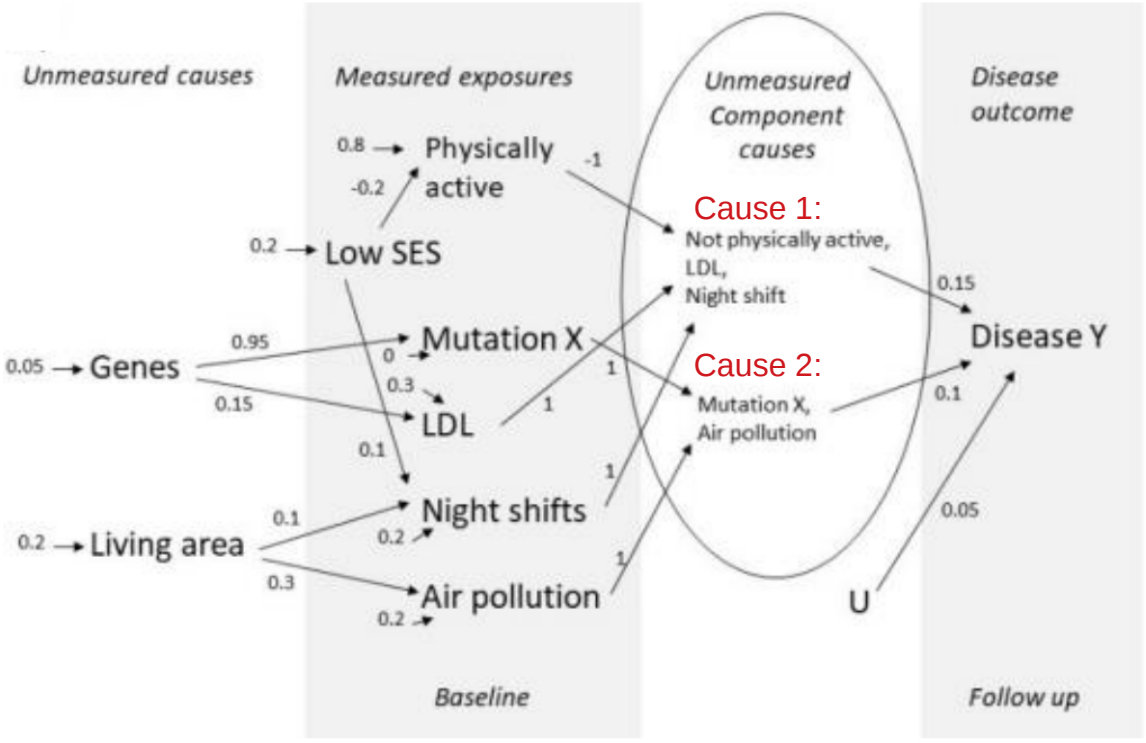
Mean risk (std) per subgroup and per exposure

Mean risk contributions by sub-group (Standard deviation) [mean risk contribution if other exposures are set to 0]	Baseline_risk	sex_0	sex_1	drug_a_0	drug_a_1	drug_b_0	drug_b_1
Sub-group 1: n=8038, e=393,Prev=80.4%, risk=4.9%, excess=3.4%, Obs risk=4.9% (4.4-5.4%) Risk based on the sum of individual effects =4.9%	4.6% (0.0%) [4.6%]						
Sub-group 2: n=950, e=194,Prev=9.5%, risk=20.3%, excess=18.8%, Obs risk=20.4% (17.9-23.2%) Risk based on the sum of individual effects =4.6%	4.6% (0.0%) [4.6%]		7.7% (0.0%) [0.0%]				7.7% (0.0%) [0.0%]
Sub-group 3: n=1012, e=208,Prev=10.1%, risk=20.4%, excess=20.2%, Obs risk=20.6% (18.1-23.2%) Risk based on the sum of individual effects =5.2%	4.6% (0.0%) [4.6%]	8% (0.1%) [0.0%]			7.7% (0.0%) [0.0%]		

If **[mean risk contrib. of exposure X_i with other exposures set to 0] < mean risk contrib of exposure X_i**

Then **synergy of X_i with other exposures in the subgroup!**

Complex simulation example



Ground truth:

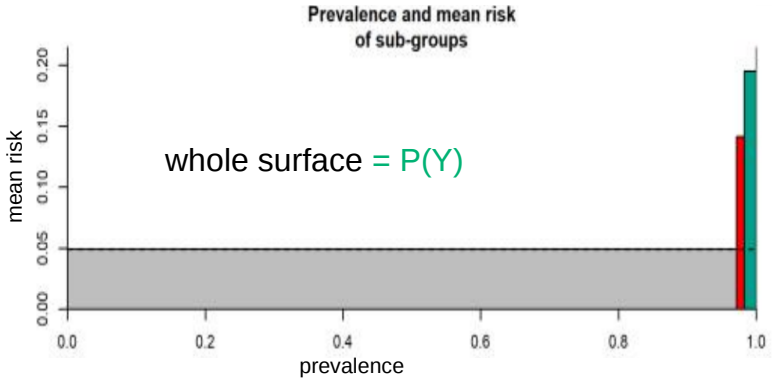
P(Y) = 5,4 % mean prediction

P(Y | U) = 5 % baseline risk

P(Cause 1) = 1,8 % prevalence
P(Y | Cause 1) = 15 % increased risk
P(Y, Cause 1)/P(Y) = 4,9 % excess fraction

P(Cause 2) = 1,2 %
P(Y | Cause 2) = 10 %
P(Y, Cause 2)/P(Y) = 2,3 %

CoOL results



whole surface = P(Y)

$\sum_i R_{X_i}$ = increased risk

R_b = baseline risk

Ground truth:

P(Y | U) = 5 % baseline risk

P(Cause 1) = 1,8 % prevalence
P(Y | Cause 1) = 15 % increased risk
P(Y, Cause 1)/P(Y) = 4,9 % excess fraction

P(Cause 2) = 1,2 %
P(Y | Cause 2) = 10 %
P(Y, Cause 2)/P(Y) = 2,3 %

F) Mean risk contributions by sub-group
(Standard diviation)
[mean risk contribution if other exposures are set to 0]

	Baseline_risk	Air_pollution_1	Air_pollution_0	Night_shifts_1	Night_shifts_0	LDL_1	LDL_0	Mutation_X_1	Mutation_X_0	Low_SES_1	Low_SES_0	Physically_active_1	Physically_active_0
Sub-group 1: n=48564, e=2378,Prev=97.1%, risk=4.9%, excess=1.2%, Obs risk=4.9% (4.7-5.1%) Risk based on the sum of individual effects =4.9%	4.9% (0.0%) [4.9%]												
Sub-group 2: n=560, e=89,Prev=1.1%, risk=14.1%, excess=2.0%, Obs risk=15.9% (13.0-19.2%) Risk based on the sum of individual effects =4.9%	4.9% (0.0%) [4.9%]	4.1% (0.1%) [0.0%]						4.4% (0.2%) [0.0%]					
Sub-group 3: n=876, e=183,Prev=1.8%, risk=19.5%, excess=4.8%, Obs risk=20.9% (18.3-23.8%) Risk based on the sum of individual effects =4.9%	4.9% (0.0%) [4.9%]			4.8% (0.0%) [0.0%]		4.8% (0.0%) [0.0%]							4.8% (0.0%) [0.0%]

BIFOLD

Fraunhofer HHI

More information about CoOL

- Tutorial and demo see project page:
  https://www.causesofoutcomelearning.org

- Open source R package to reproduce results (including plots):
  https://cran.r-project.org/package=CoOL

- Supplementary material of the paper (including various controlled simulations, robustness checks and a real-world example):
  https://doi.org/10.1093/ije/dyac078