Valid Causal Inference with (Some) Invalid Instruments

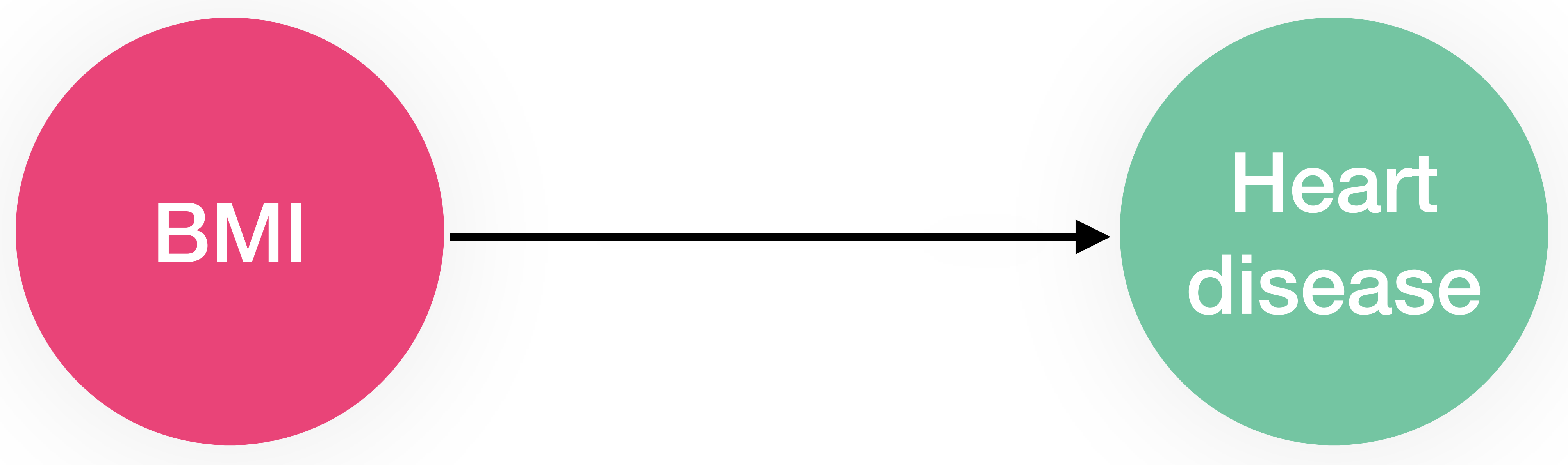
Jason Hartford, Victor Veitch, Dhanya Sridhar, Kevin Leyton-Brown

THE UNIVERSITY OF BRITISH COLUMBIA

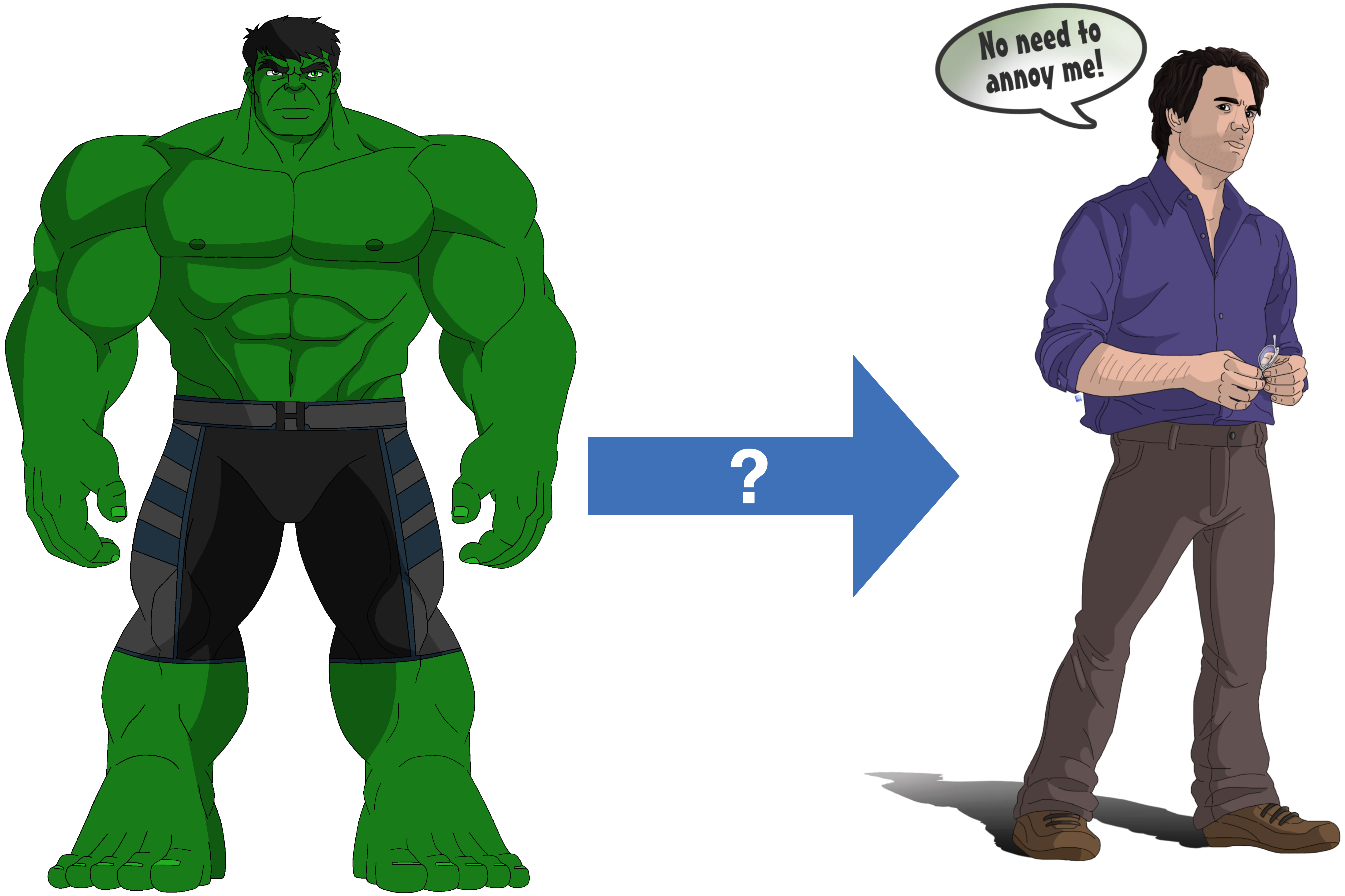
COLUMBIA UNIVERSITY
DATA SCIENCE INSTITUTE

THE UNIVERSITY OF CHICAGO

Causal inference with unobserved confounding



Causal inference with unobserved confounding



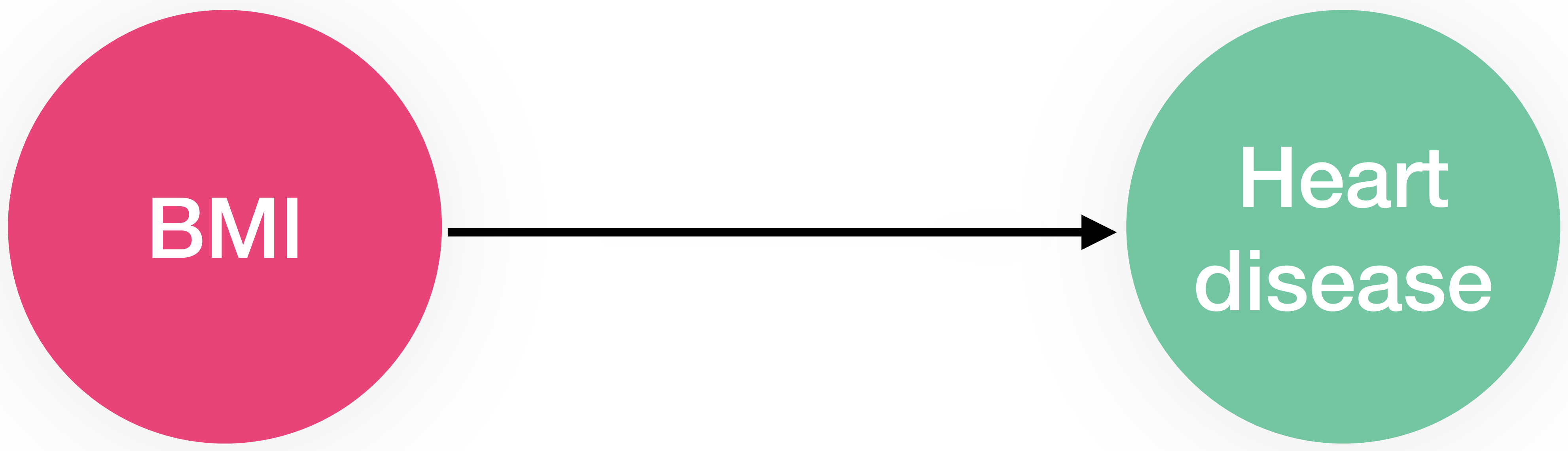
BMI: 50

P(Heart Disease | do(BMI = 50))

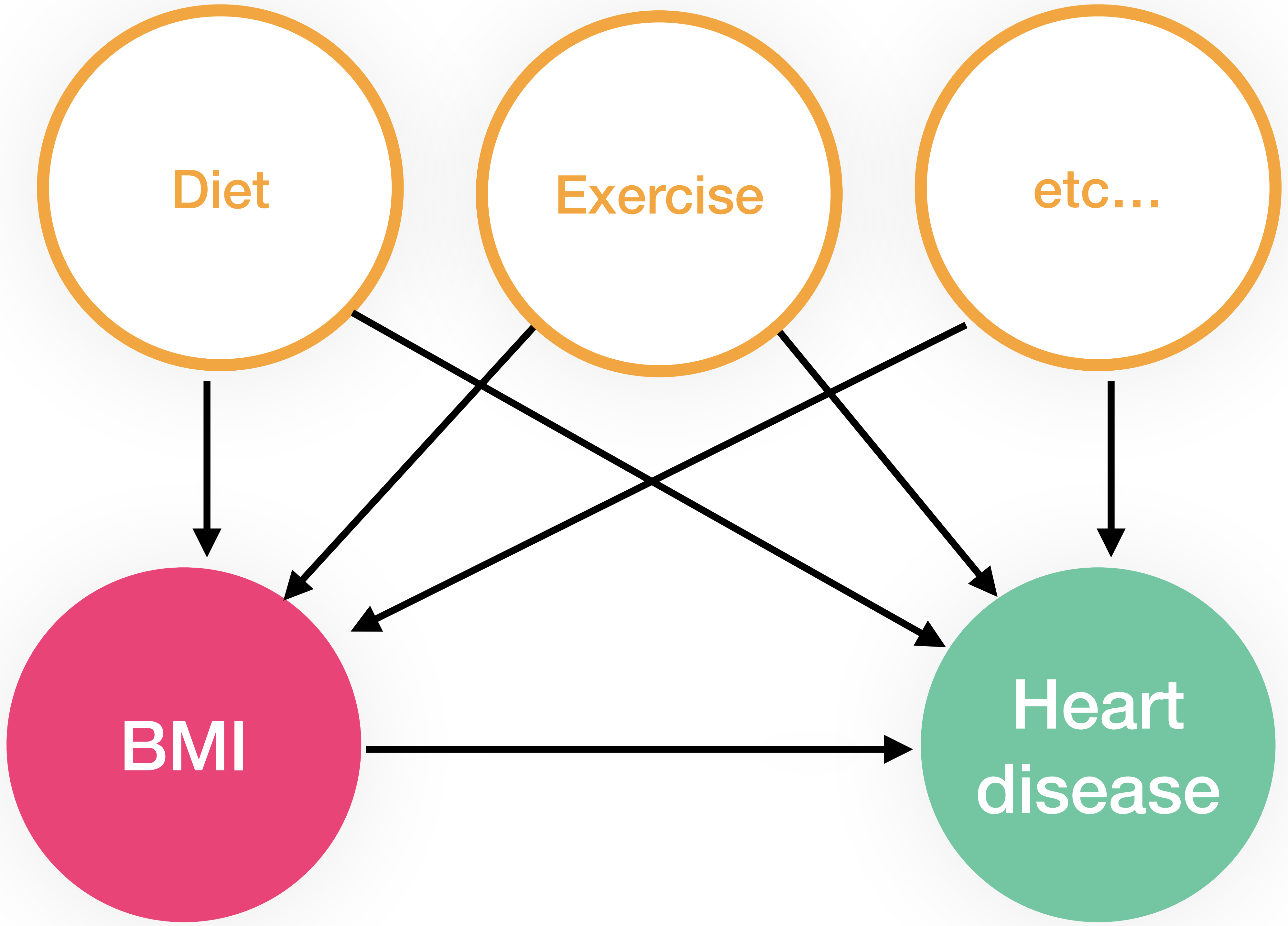
BMI: 23

P(Heart Disease | do(BMI=23))

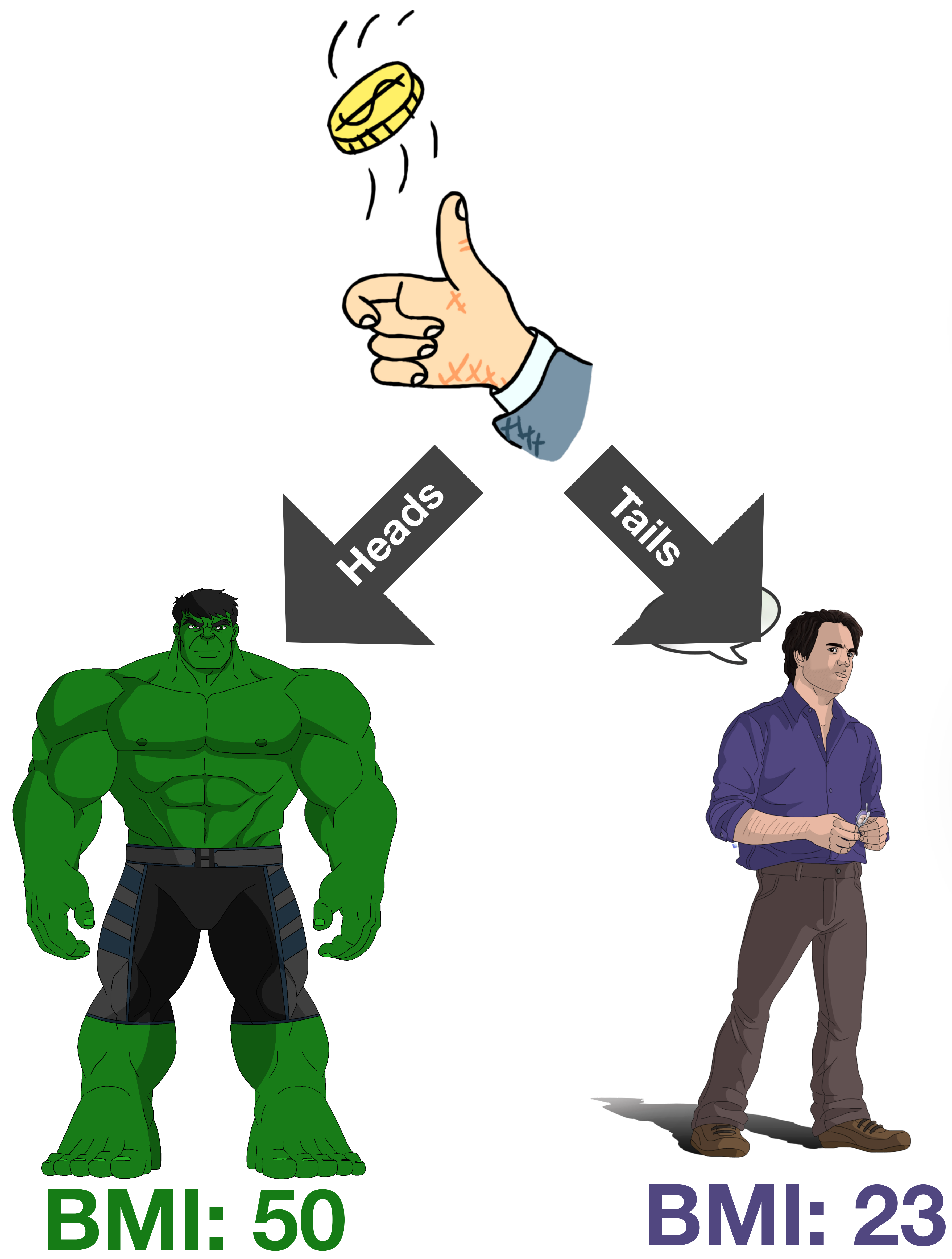
Causal inference with unobserved confounding

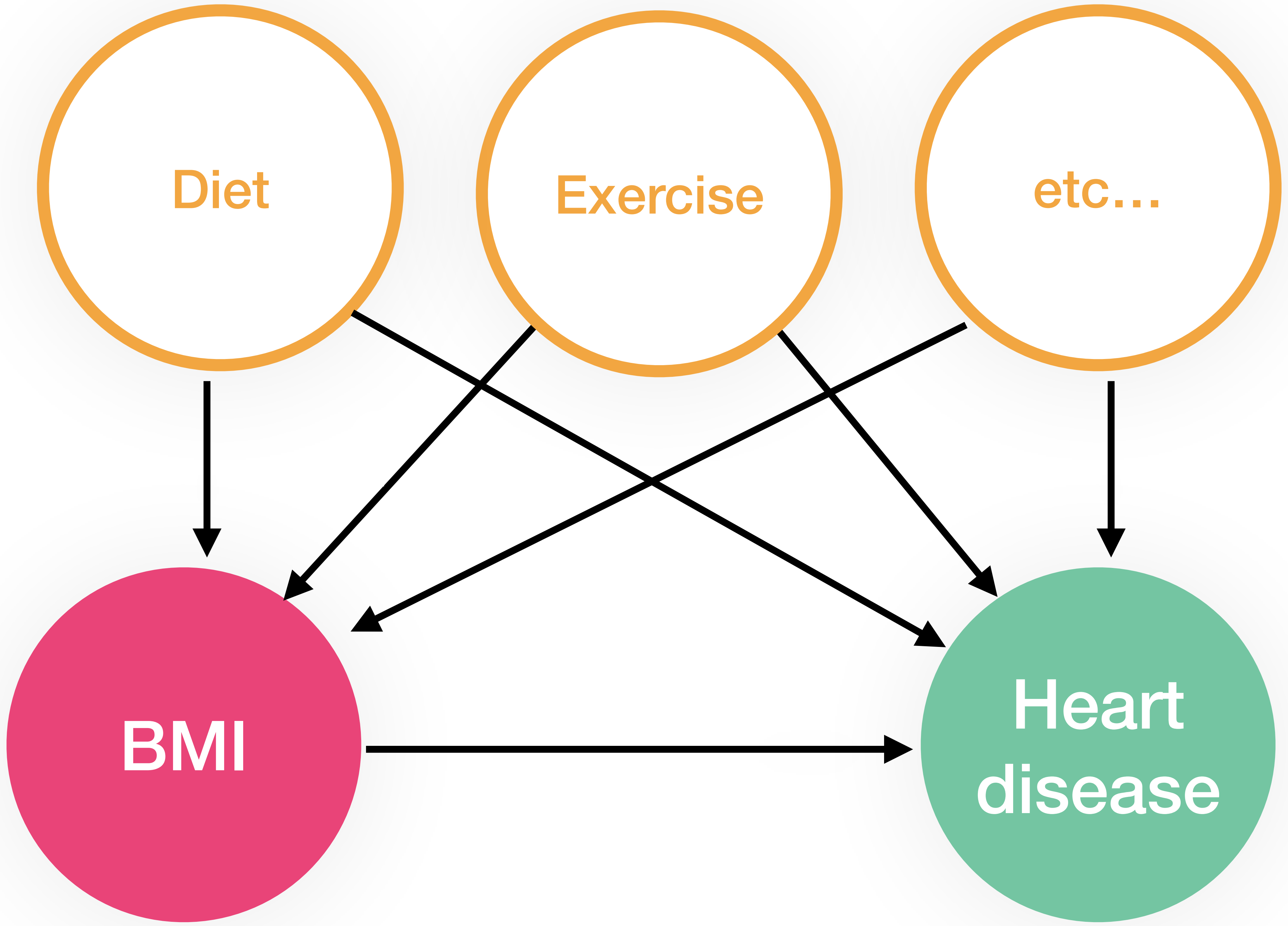


Causal inference with unobserved confounding

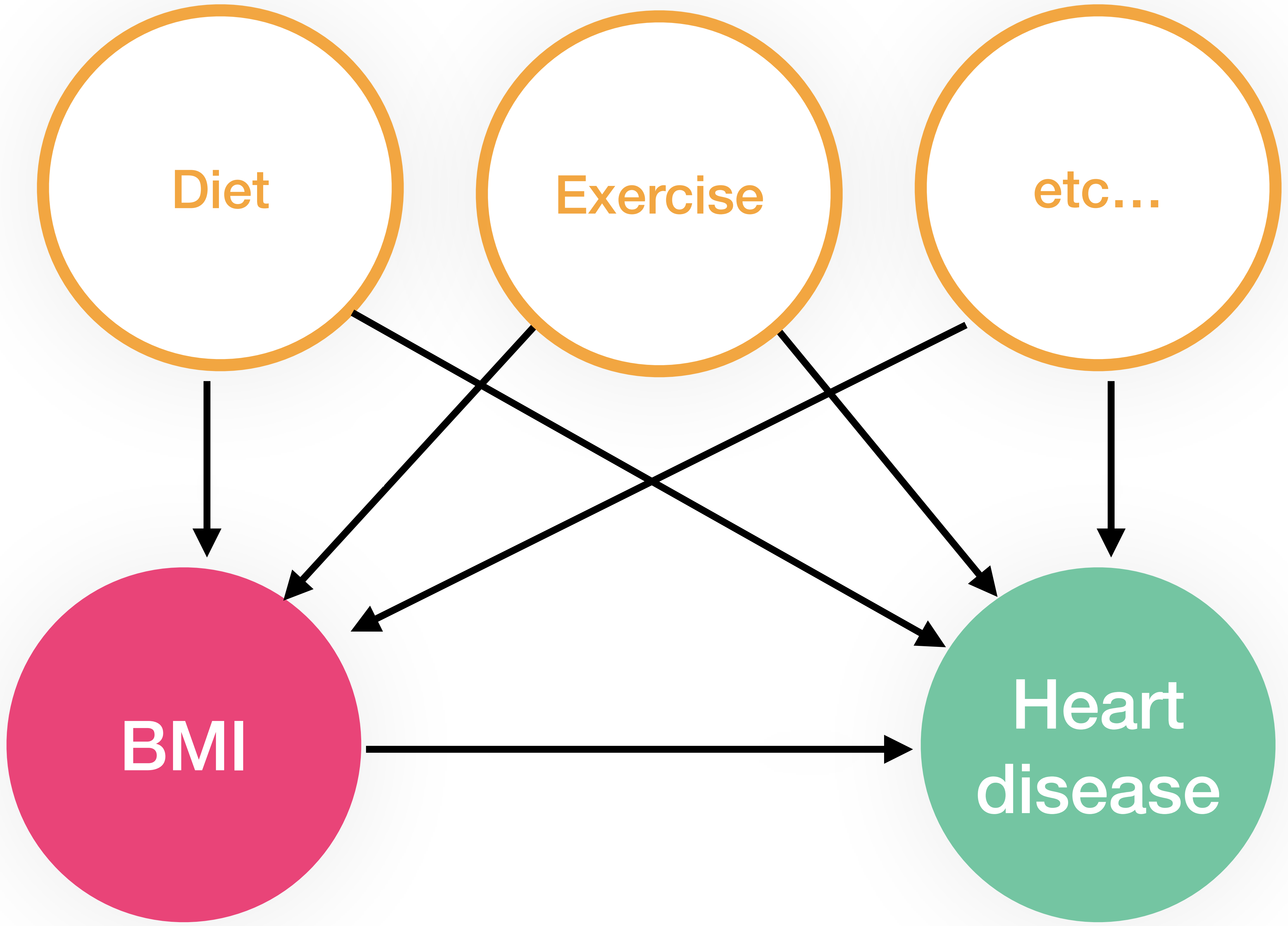


Causal inference with unobserved confounding

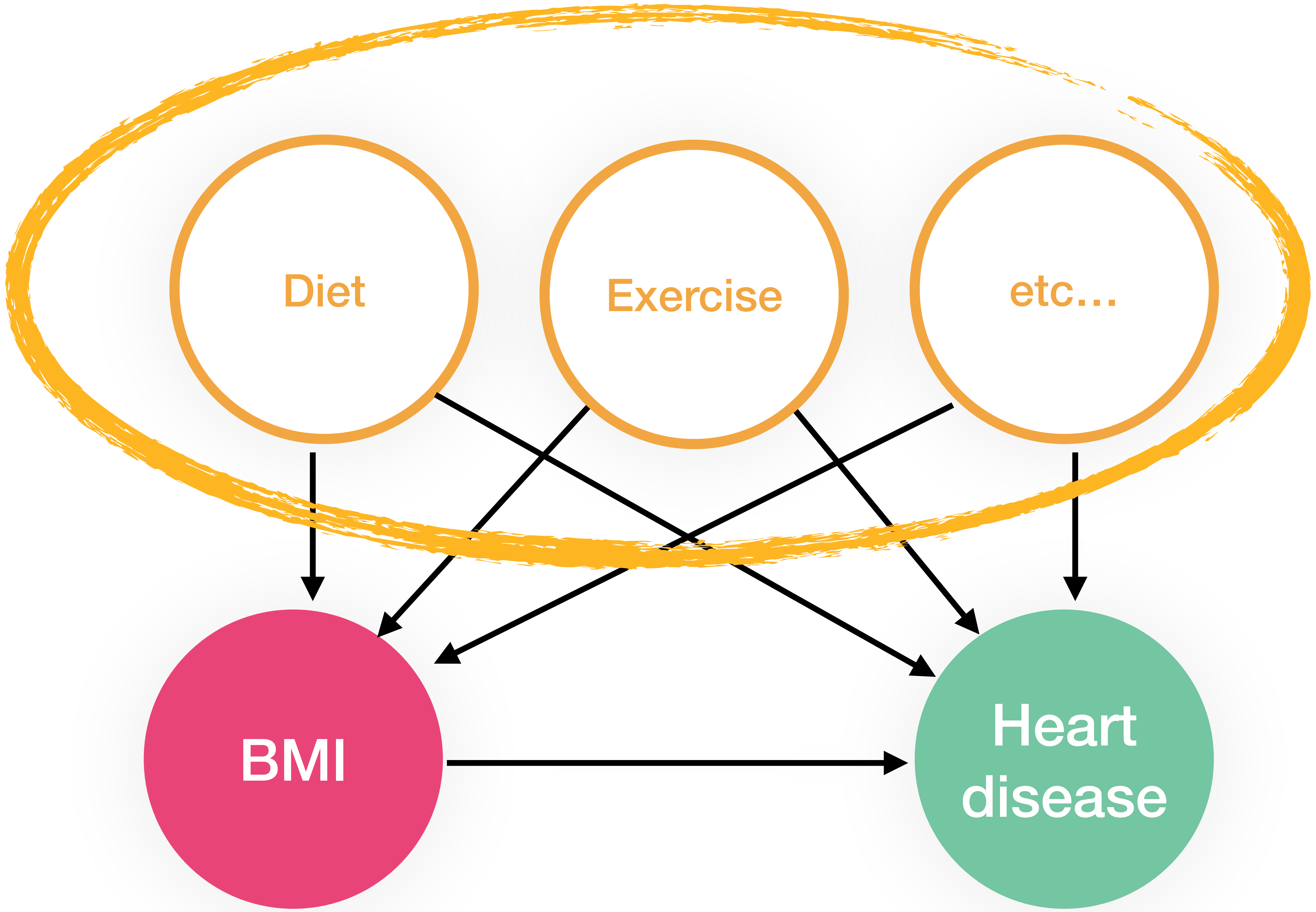




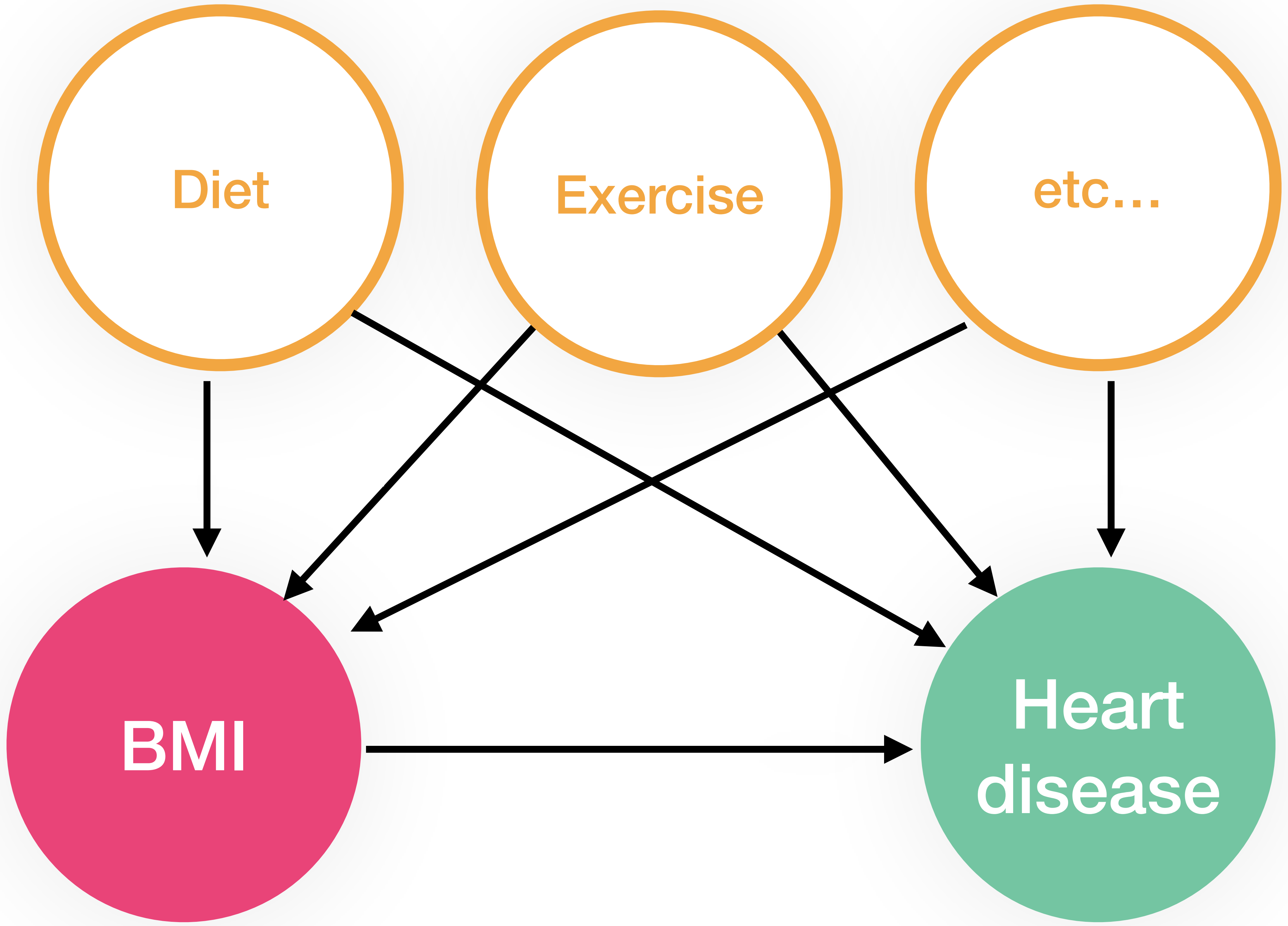
Causal inference with unobserved confounding



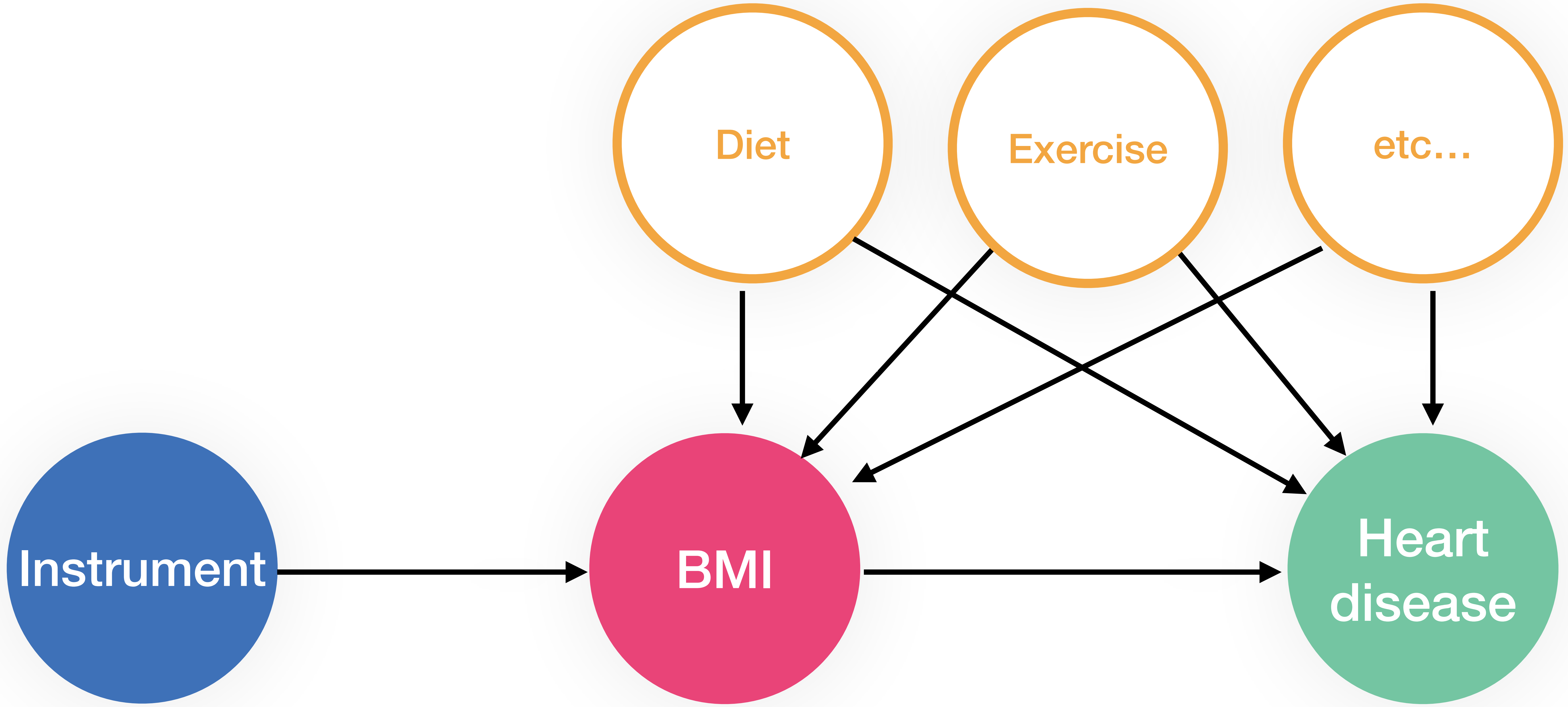
Causal inference with unobserved confounding



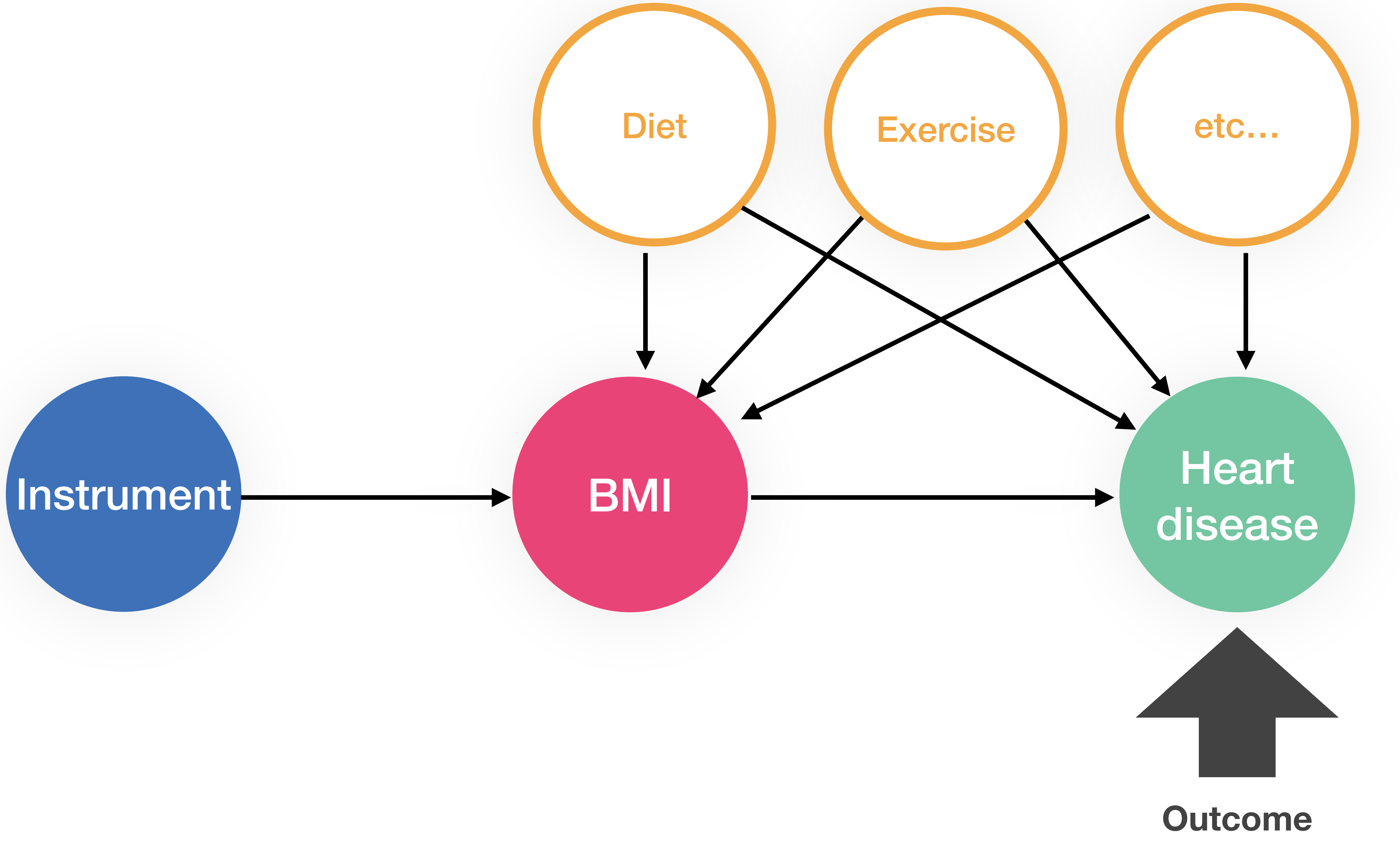
Causal inference with unobserved confounding



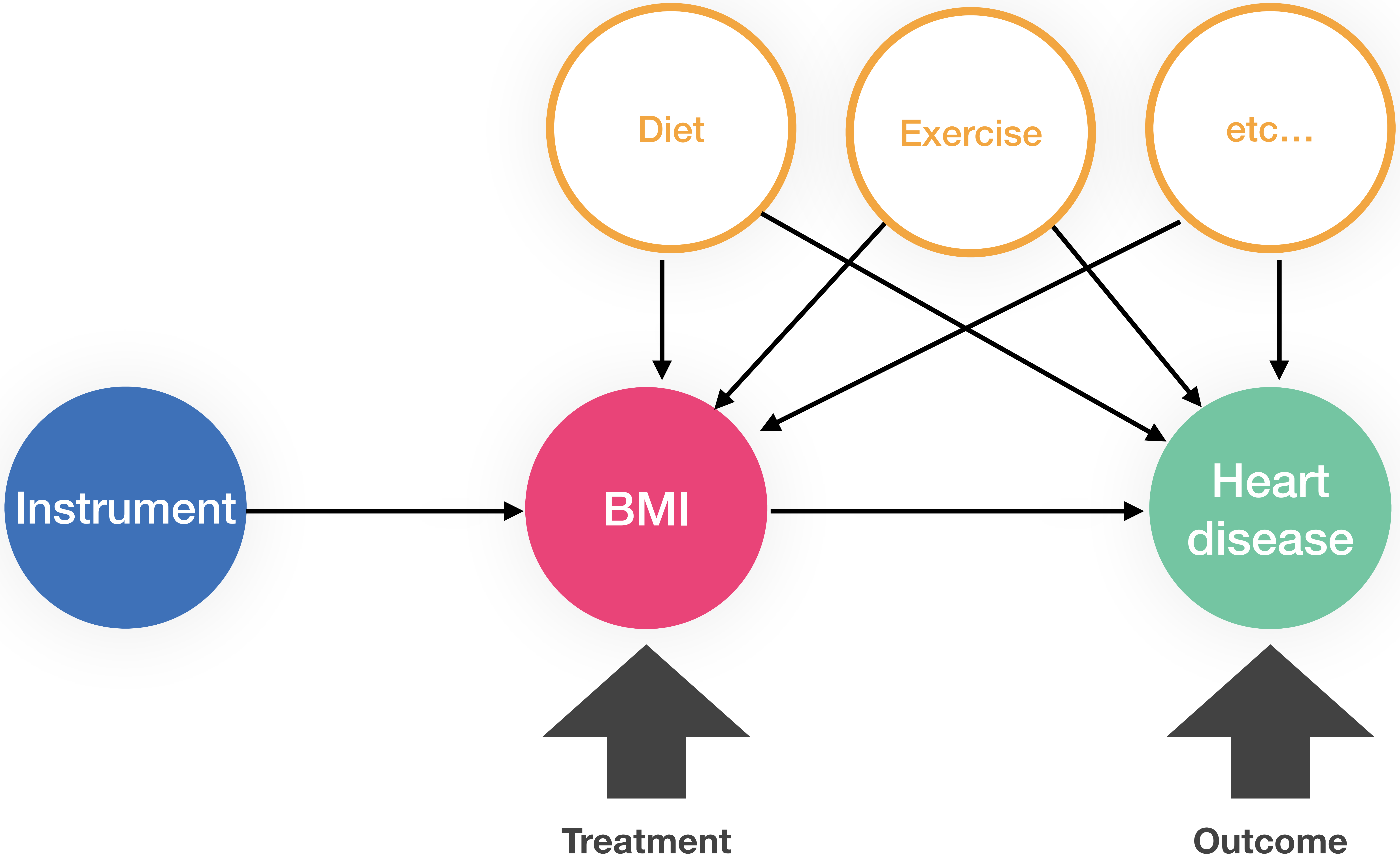
Causal inference with unobserved confounding



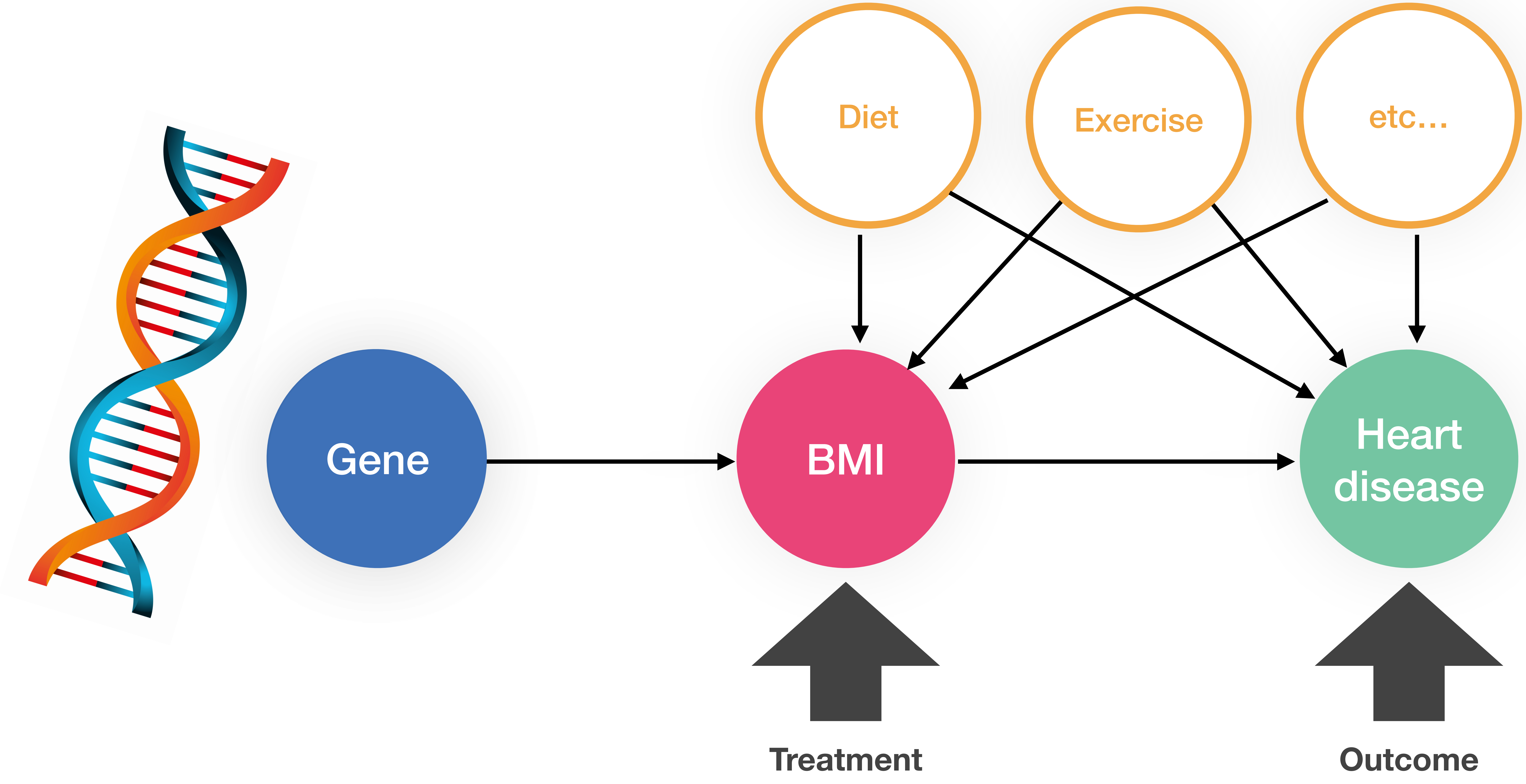
Causal inference with unobserved confounding



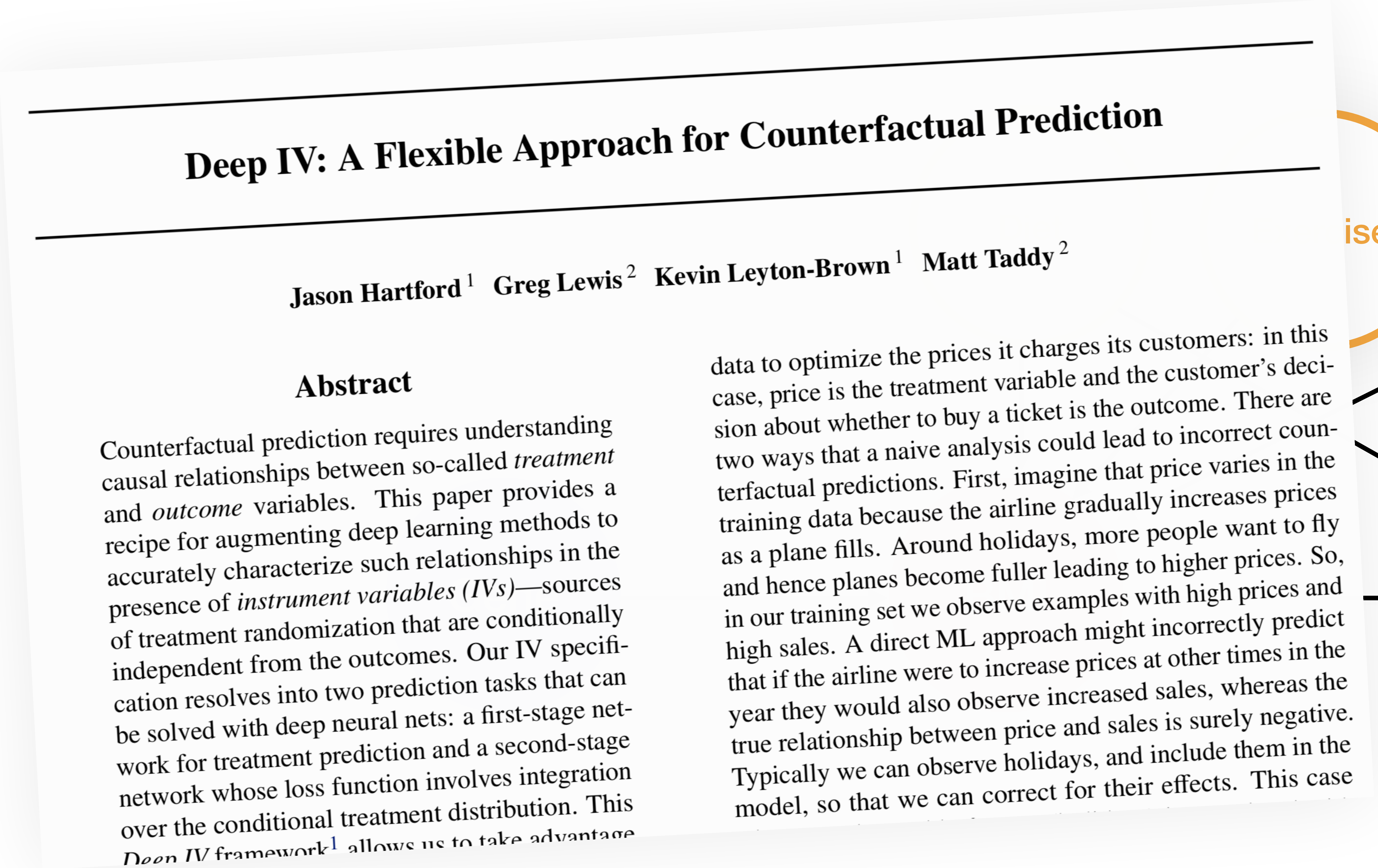
Causal inference with unobserved confounding



Causal inference with unobserved confounding



Causal inference with unobserved confounding



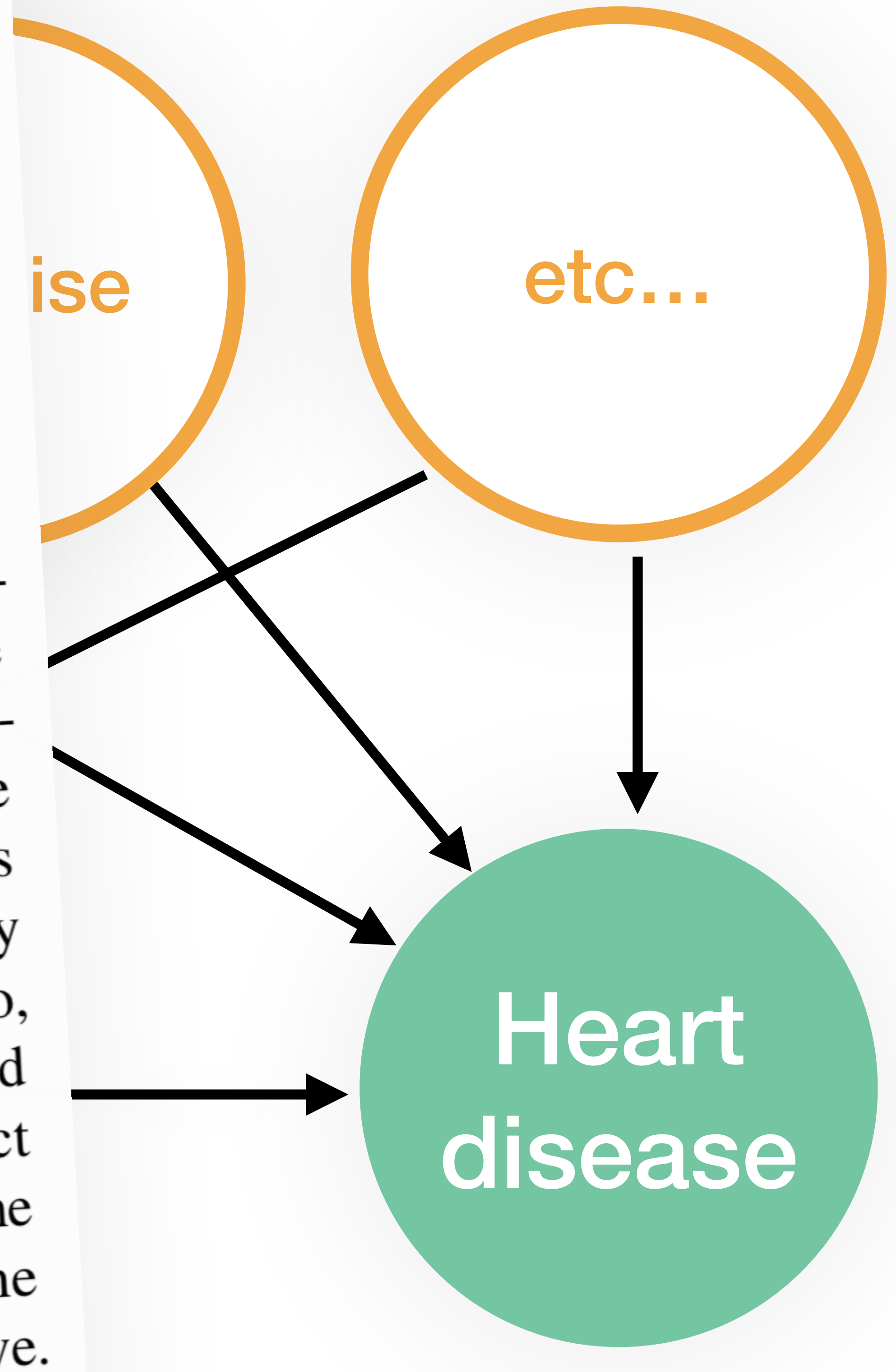
Deep IV: A Flexible Approach for Counterfactual Prediction

Jason Hartford[1] Greg Lewis[2] Kevin Leyton-Brown[1] Matt Taddy[2]

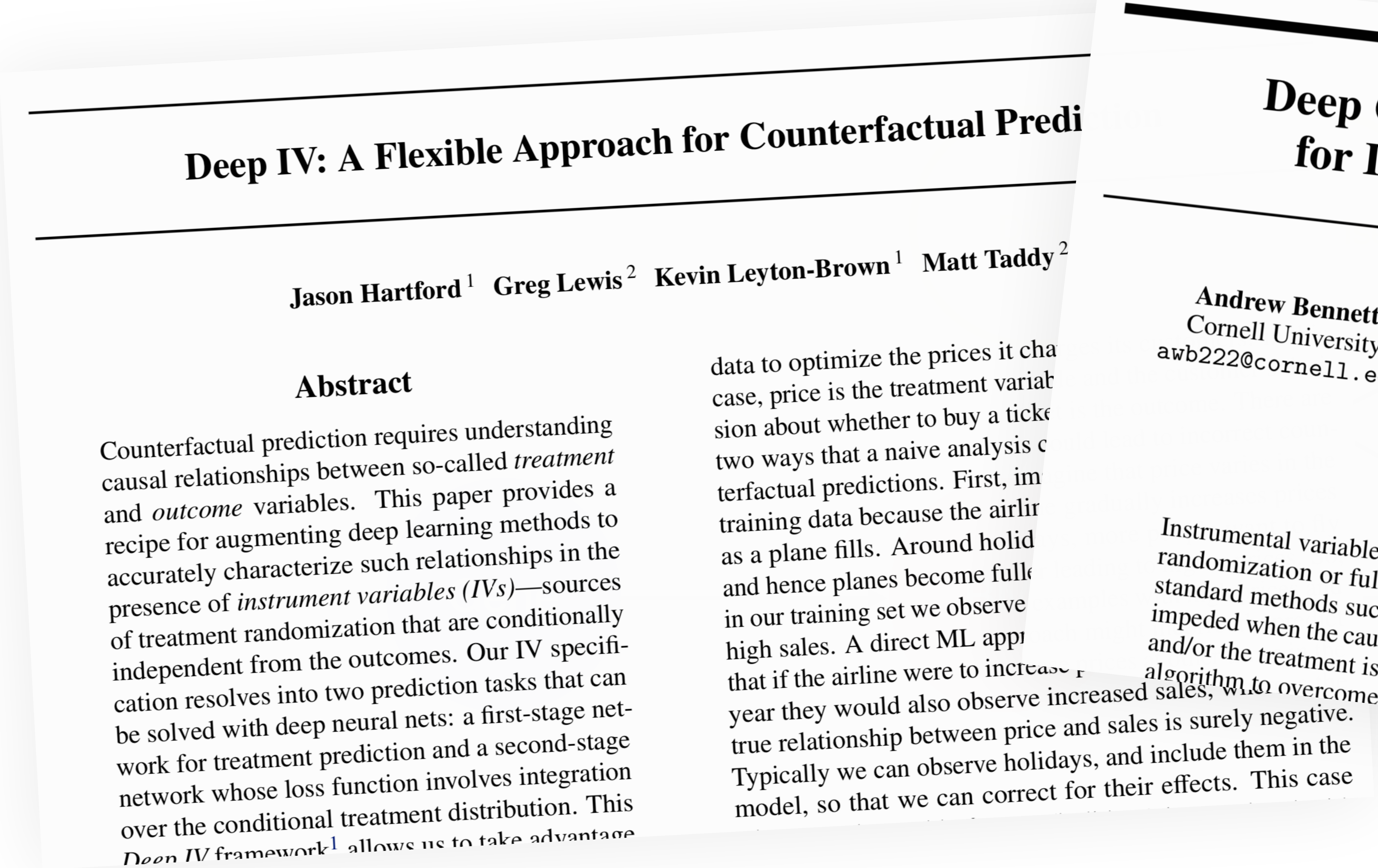
Abstract

Counterfactual prediction requires understanding causal relationships between so-called *treatment* and *outcome* variables. This paper provides a recipe for augmenting deep learning methods to accurately characterize such relationships in the presence of *instrument variables (IVs)*—sources of treatment randomization that are conditionally independent from the outcomes. Our IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This *Deep IV* framework[1] allows us to take advantage

data to optimize the prices it charges its customers: in this case, price is the treatment variable and the customer's decision about whether to buy a ticket is the outcome. There are two ways that a naive analysis could lead to incorrect counterfactual predictions. First, imagine that price varies in the training data because the airline gradually increases prices as a plane fills. Around holidays, more people want to fly and hence planes become fuller leading to higher prices. So, in our training set we observe examples with high prices and high sales. A direct ML approach might incorrectly predict that if the airline were to increase prices at other times in the year they would also observe increased sales, whereas the true relationship between price and sales is surely negative. Typically we can observe holidays, and include them in the model, so that we can correct for their effects. This case



Causal inference with unobserved confounding

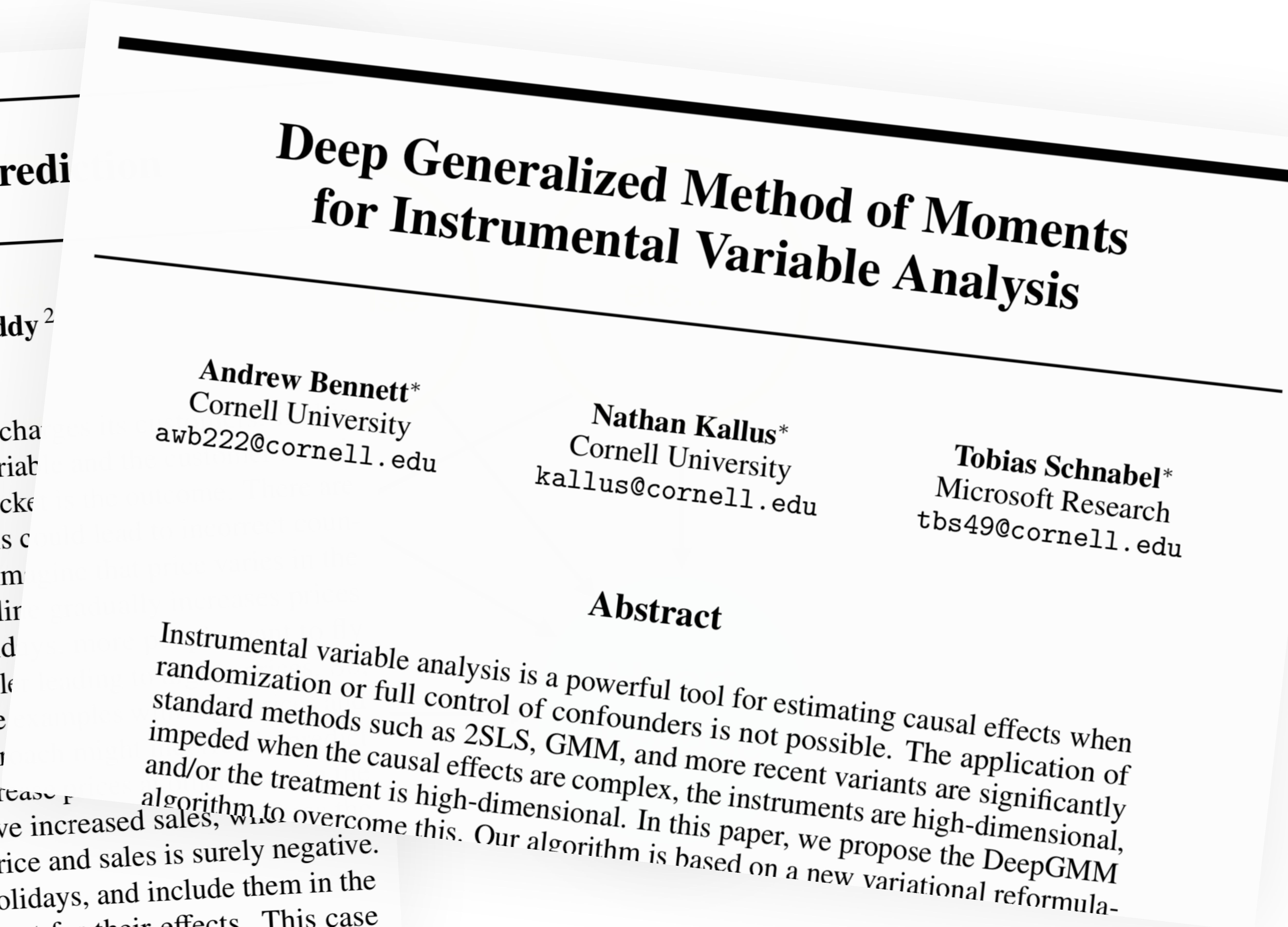

Deep IV: A Flexible Approach for Counterfactual Predi

Jason Hartford [1] Greg Lewis [2] Kevin Leyton-Brown [1] Matt Taddy [2]

Abstract

Counterfactual prediction requires understanding causal relationships between so-called *treatment* and *outcome* variables. This paper provides a recipe for augmenting deep learning methods to accurately characterize such relationships in the presence of *instrument variables (IVs)*—sources of treatment randomization that are conditionally independent from the outcomes. Our IV specification resolves into two prediction tasks that can be solved with deep neural nets: a first-stage network for treatment prediction and a second-stage network whose loss function involves integration over the conditional treatment distribution. This *Deep IV* framework[1] allows us to take advantage

data to optimize the prices it cha
case, price is the treatment variab
sion about whether to buy a ticke
two ways that a naive analysis c
terfactual predictions. First, im
training data because the airlir
as a plane fills. Around holid
and hence planes become fulle
in our training set we observe
high sales. A direct ML appr
that if the airline were to increase
year they would also observe increased sales,
true relationship between price and sales is surely negative.
Typically we can observe holidays, and include them in the
model, so that we can correct for their effects. This case


Deep Generalized Method of Moments for Instrumental Variable Analysis

Andrew Bennett*
Cornell University
awb222@cornell.edu

Nathan Kallus*
Cornell University
kallus@cornell.edu

Tobias Schnabel*
Microsoft Research
tbs49@cornell.edu

Abstract

Instrumental variable analysis is a powerful tool for estimating causal effects when randomization or full control of confounders is not possible. The application of standard methods such as 2SLS, GMM, and more recent variants are significantly impeded when the causal effects are complex, the instruments are high-dimensional, and/or the treatment is high-dimensional. In this paper, we propose the DeepGMM algorithm to overcome this. Our algorithm is based on a new variational reformula-

Causal inference with unobserved confounding

Deep IV: A Flexible Approach for Counterfactual Predi

Counterfactual p
causal relationsh
and *outcome* va
recipe for augm
accurately chara
presence of *inst*
of treatment ran
independent fro
cation resolves i
be solved with c
work for treatm
network whose
over the conditi
Deep IV framew

Deep Generalized Method of Moments
for Instrumental Variable Analysis

t*
y
edu

Nathan Kallus*
Cornell University
kallus@cornell.edu

Tobias Schnabel*
Microsoft Research
tbs49@cornell.edu

Abstract

analysis is a powerful tool for estimating causal effects when
control of confounders is not possible. The application of
as 2SLS, GMM, and more recent variants are significantly
al effects are complex, the instruments are high-dimensional,
high-dimensional. In this paper, we propose the DeepGMM
this. Our algorithm is based on a new variational reformula-

Kernel Instrumental Variable Regression

Rahul Singh
MIT Economics
rahul.singh@mit.edu

Maneesh Sahani
Gatsby Unit, UCL
maneesh@gatsby.ucl.ac.uk

Arthur Gretton
Gatsby Unit, UCL
arthur.gretton@gmail.com

Abstract

Instrumental variable (IV) regression is a strategy for learning causal relationships in observational data. If measurements of input X and output Y are confounded, the causal relationship can nonetheless be identified if an instrumental variable Z is available that influences X directly, but is conditionally independent of Y given X and the unmeasured confounder. The classic two-stage least squares algorithm (2SLS) simplifies the estimation problem by modeling all relationships as linear functions. We propose kernel instrumental variable regression (KIV), a nonparametric generalization of 2SLS, modeling relations among X, Y, and Z as

Causal inference with unobserved confounding

Minimax Estimation of Conditional Moment Models

Nishanth Dikkala
MIT
nishanthd@csail.mit.edu

Greg Lewis
Microsoft Research
glewis@microsoft.com

Lester Mackey
Microsoft Research
lmackey@microsoft.com

Vasilis Syrgkanis
Microsoft Research
vasy@microsoft.com

Abstract

We develop an approach for estimating models described via conditional moment restrictions, with a prototypical application being non-parametric instrumental variable regression. We introduce a min-max criterion function, under which the estimation problem can be thought of as solving a zero-sum game between a modeler who is optimizing over the hypothesis space of the target model and an adversary who identifies violating moments over a test function space. We analyze the statistical estimation rate of the resulting estimator for arbitrary hypothesis spaces, with respect to an appropriate analogue of the mean squared error metric, for ill-posed inverse problems. We show that when the minimax criterion is regularized with a second moment penalty on the test function and the test function space is sufficiently rich, then the estimation rate scales with the critical radius

Causal inference with unobserved confounding

Learning Deep Features in Instrumental Variable Regression

Liyuan Xu
Gatsby Unit
liyuan.jo.19@ucl.ac.uk

Yutian Chen
DeepMind
yutianc@google.com

Siddarth Srinivasan
University of Washington
sidsrini@cs.washington.edu

Nando de Freitas
DeepMind
nandodefreitas@google.com

Arnaud Doucet
DeepMind
arnauddoucet@google.com

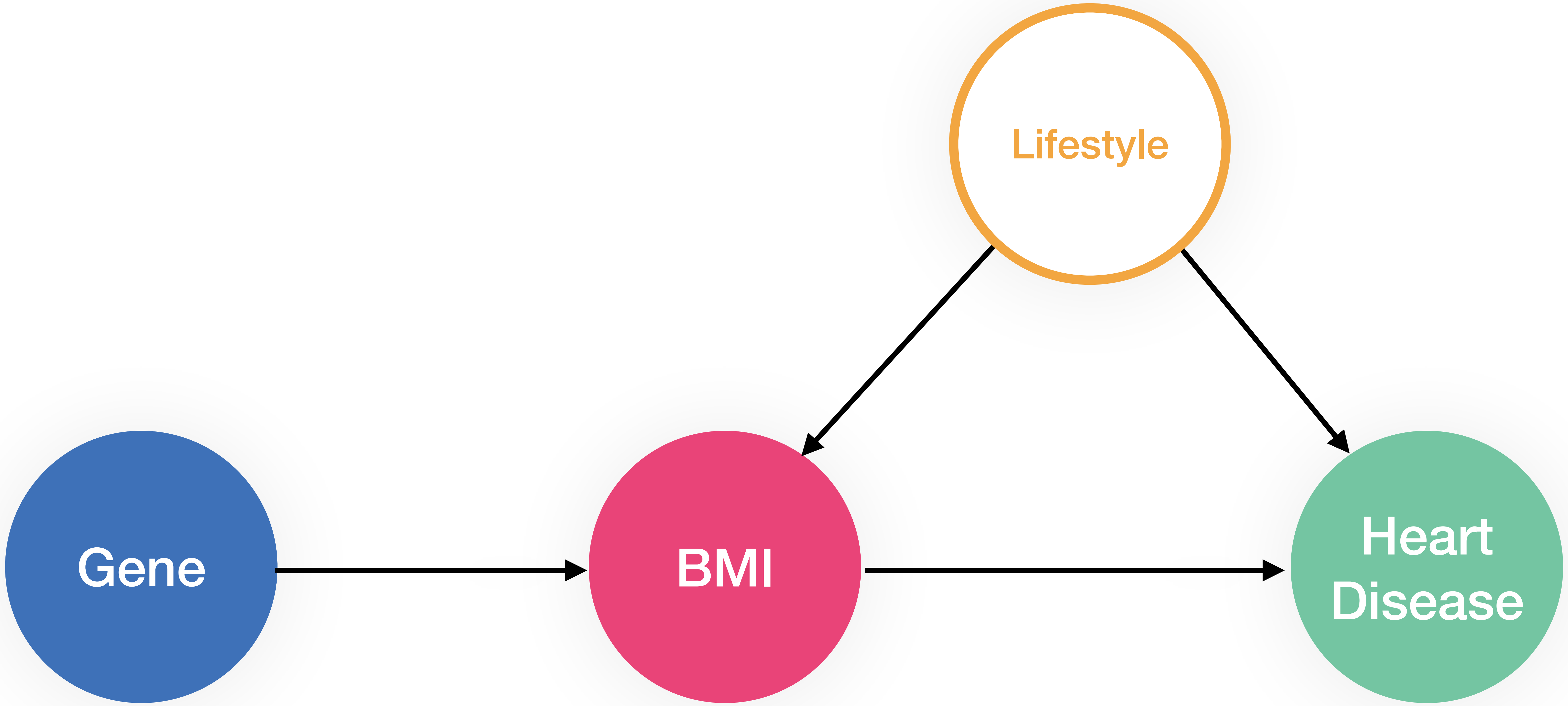
Arthur Gretton
Gatsby Unit
arthur.gretton@gmail.com

November 3, 2020

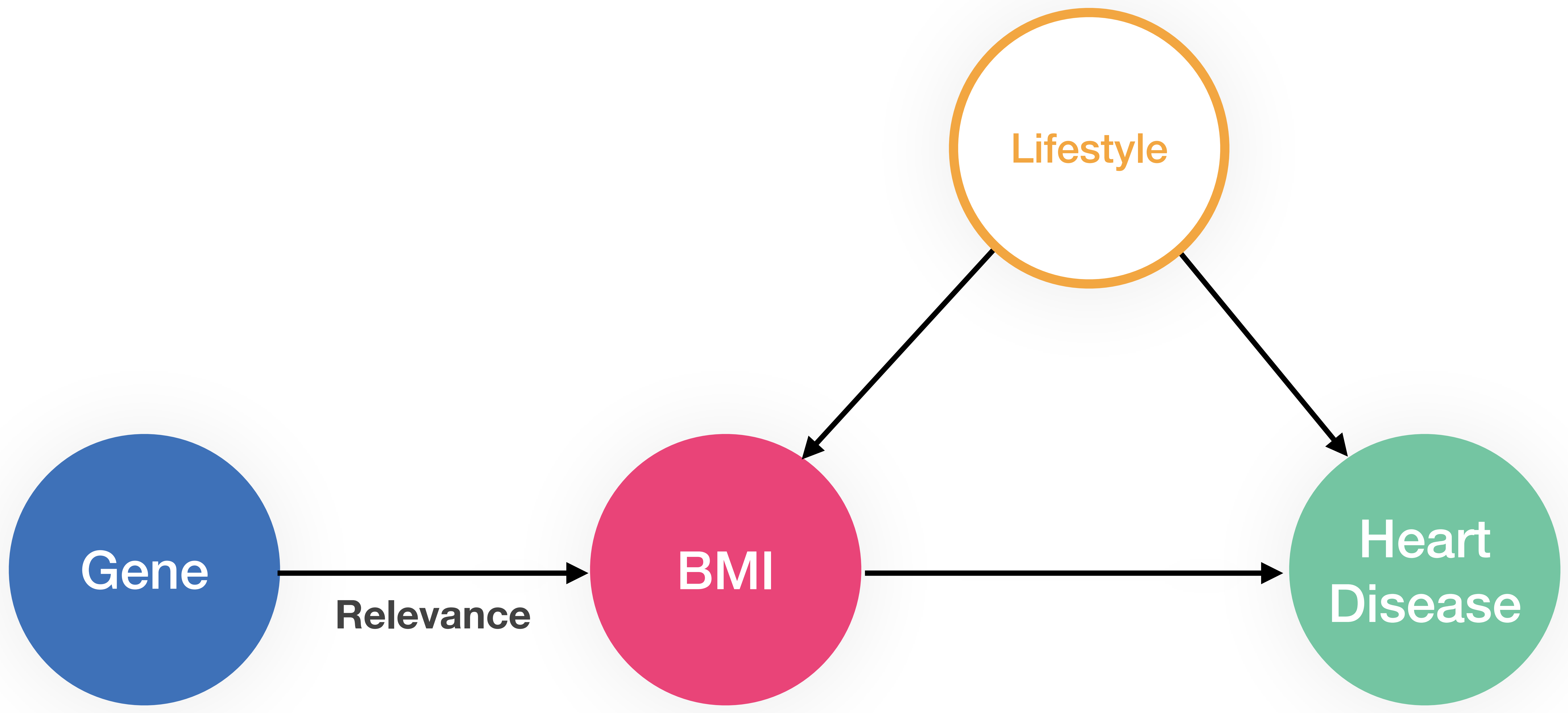
Abstract

Instrumental variable (IV) regression is a standard strategy for learning causal relationships between confounded treatment and outcome variables from observational data by utilizing an instrumental variable, which affects the outcome only through the treatment. In classical IV regression, learning proceeds in two stages: stage 1 performs linear regression from the instrument to the treatment; and stage 2 performs linear regression from the treatment to the outcome, conditioned on the instrument. We propose a novel method, *deep feature instrumental variable regression (DFIV)*, to address the case where relations between instruments, treatments, and outcomes may be nonlinear. In this case, deep neural nets are trained to define informative nonlinear features on the instruments and treatments. We propose an alternating training regime for these features to ensure good end-to-end performance when composing stages 1 and 2, thus obtaining highly flexible feature maps in a computationally efficient manner. DFIV outperforms recent state-of-the-art methods on challenging IV benchmarks, including settings involving high dimensional

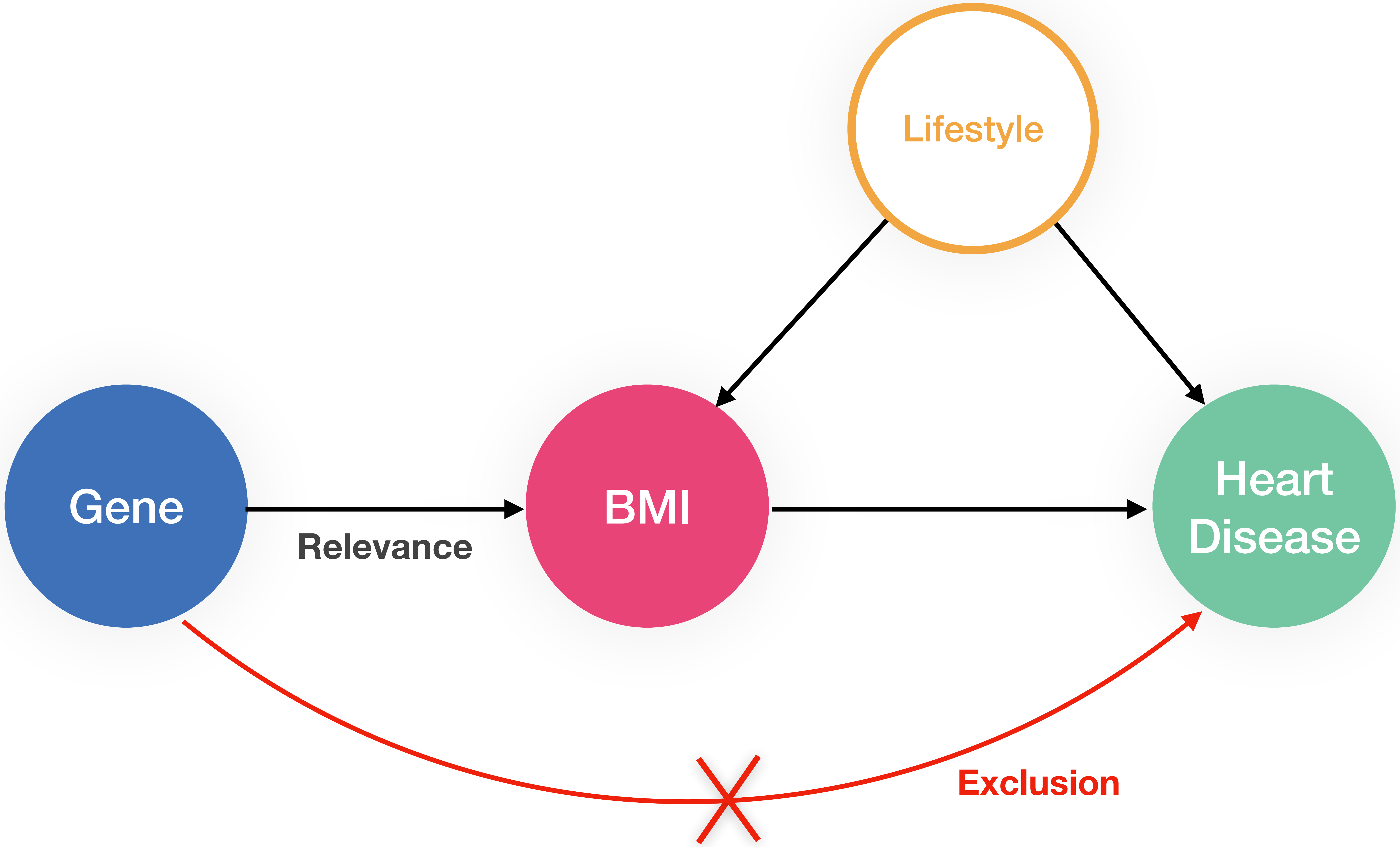
Instrumental variable assumptions



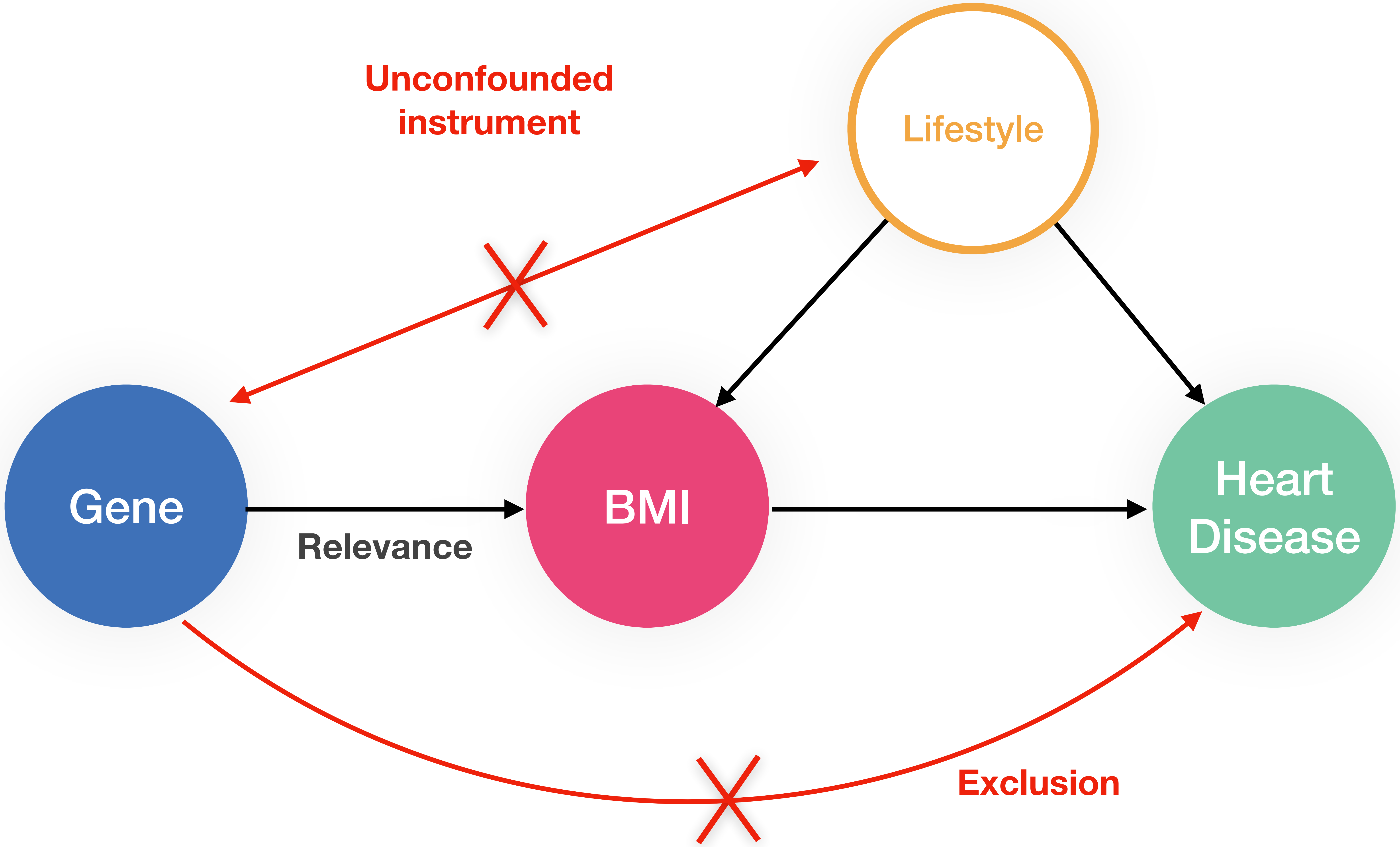
Instrumental variable assumptions



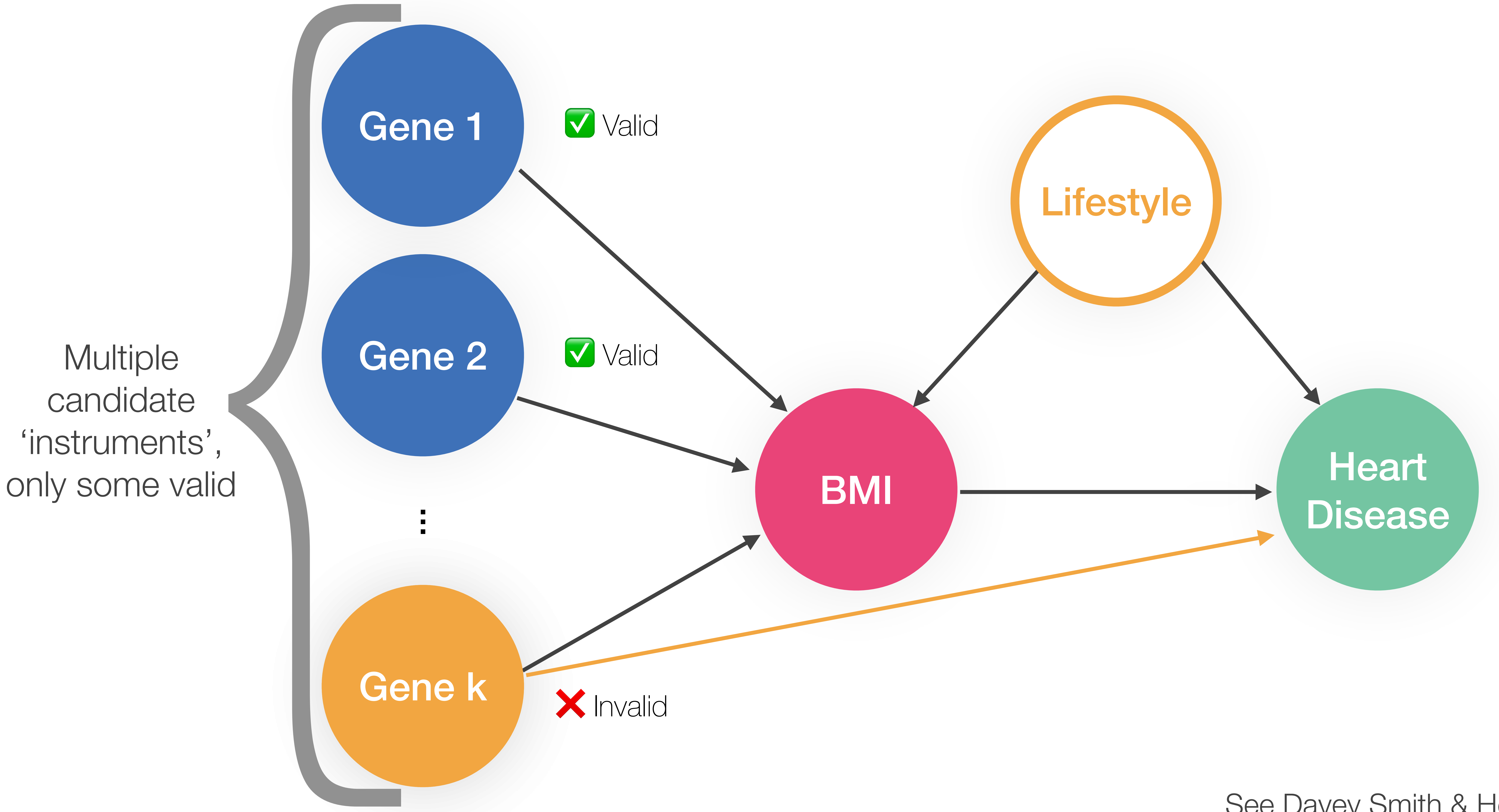
Instrumental variable assumptions



Instrumental variable assumptions

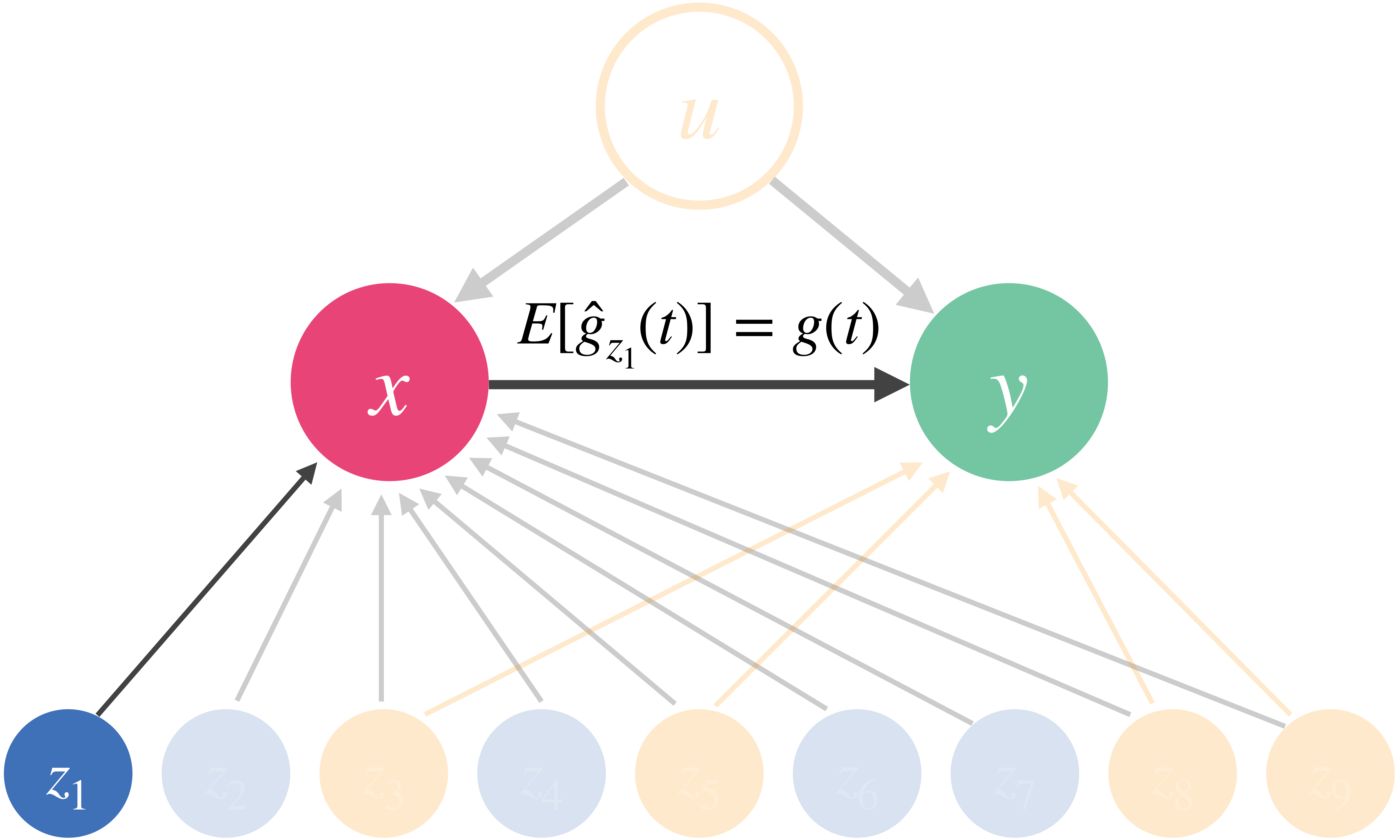


Example: Mendelian randomization



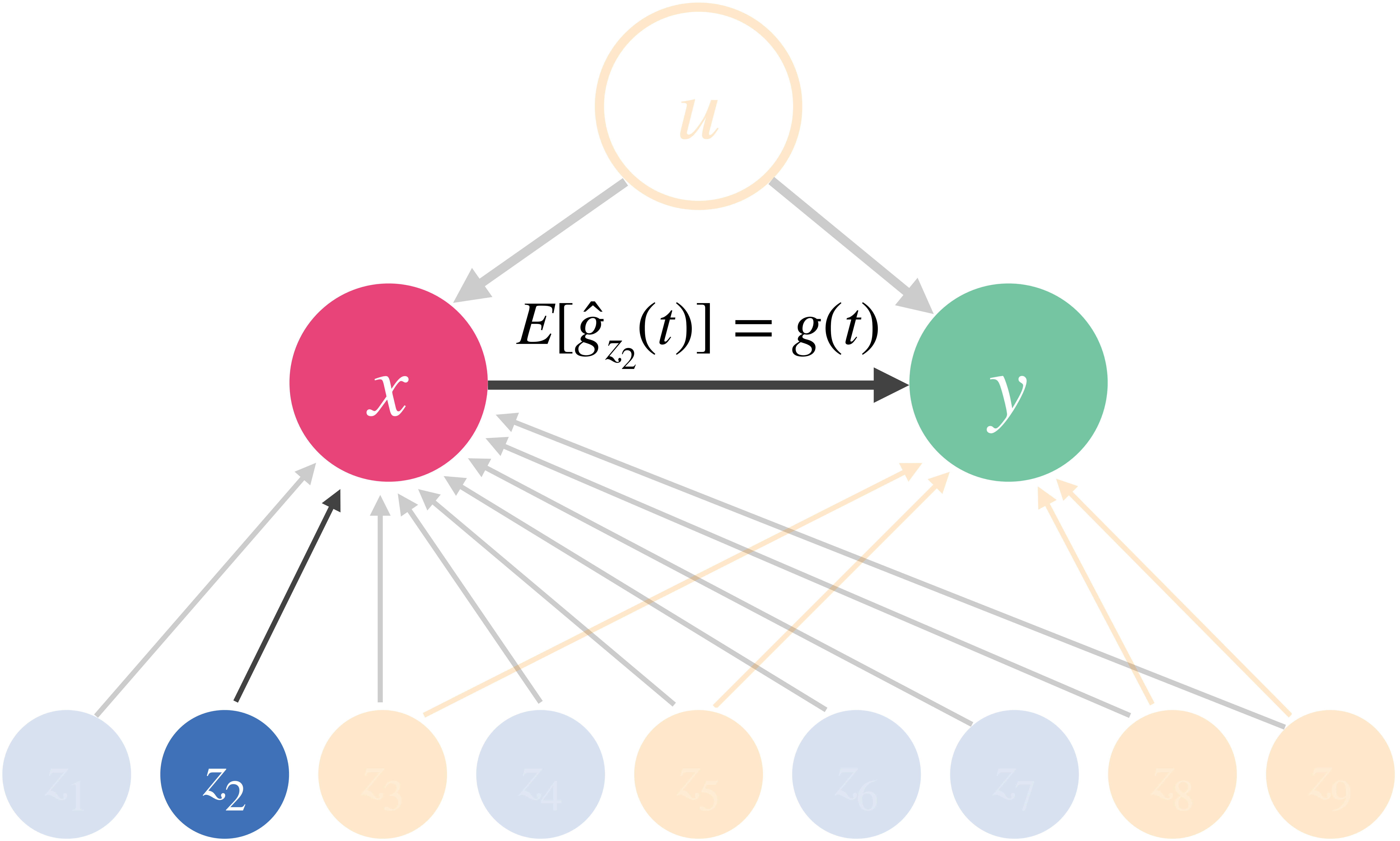
See Davey Smith & Hemani 2014 for a review

One way to be unbiased



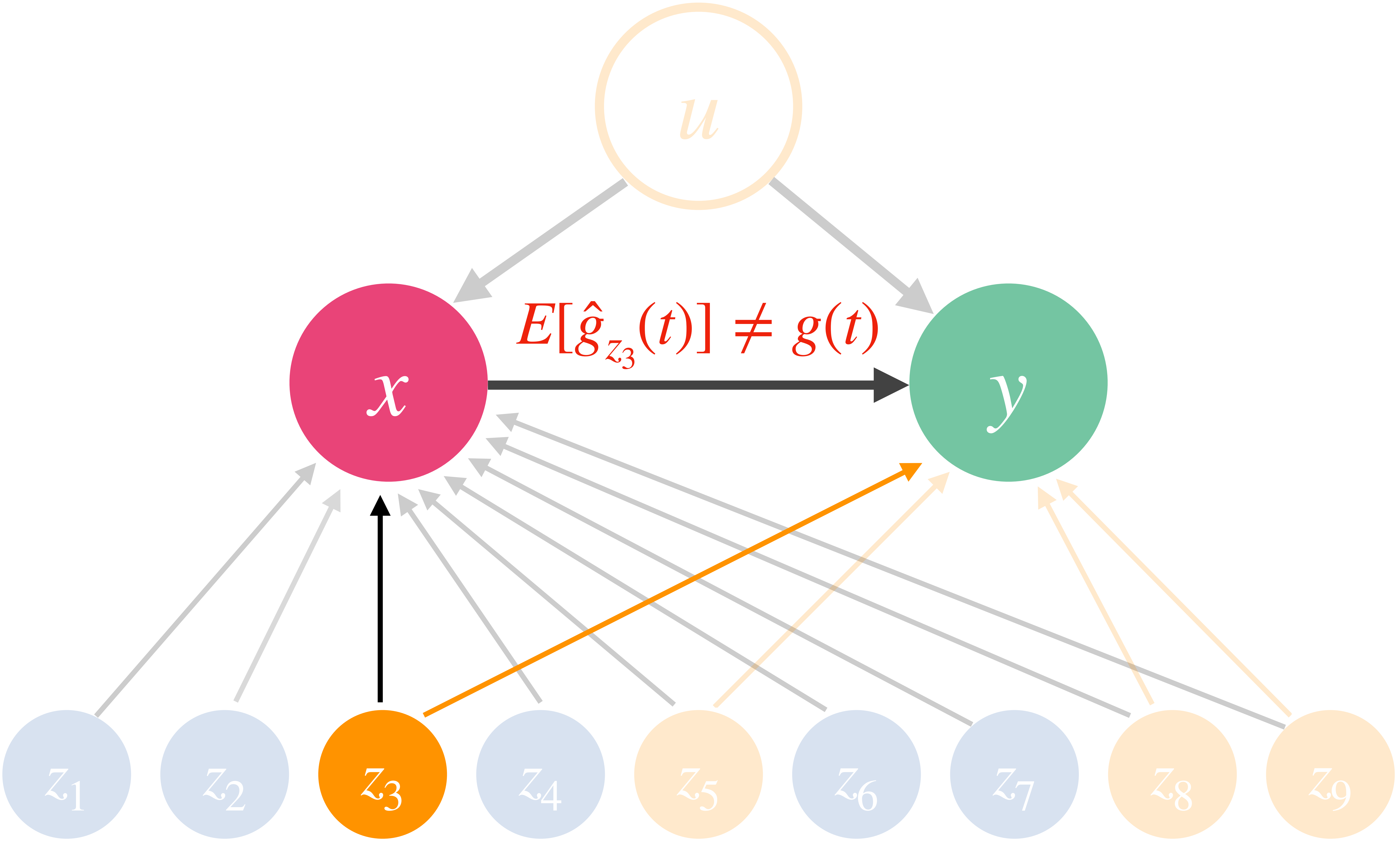
[c.f. Hartwig et al. 2017]

One way to be unbiased



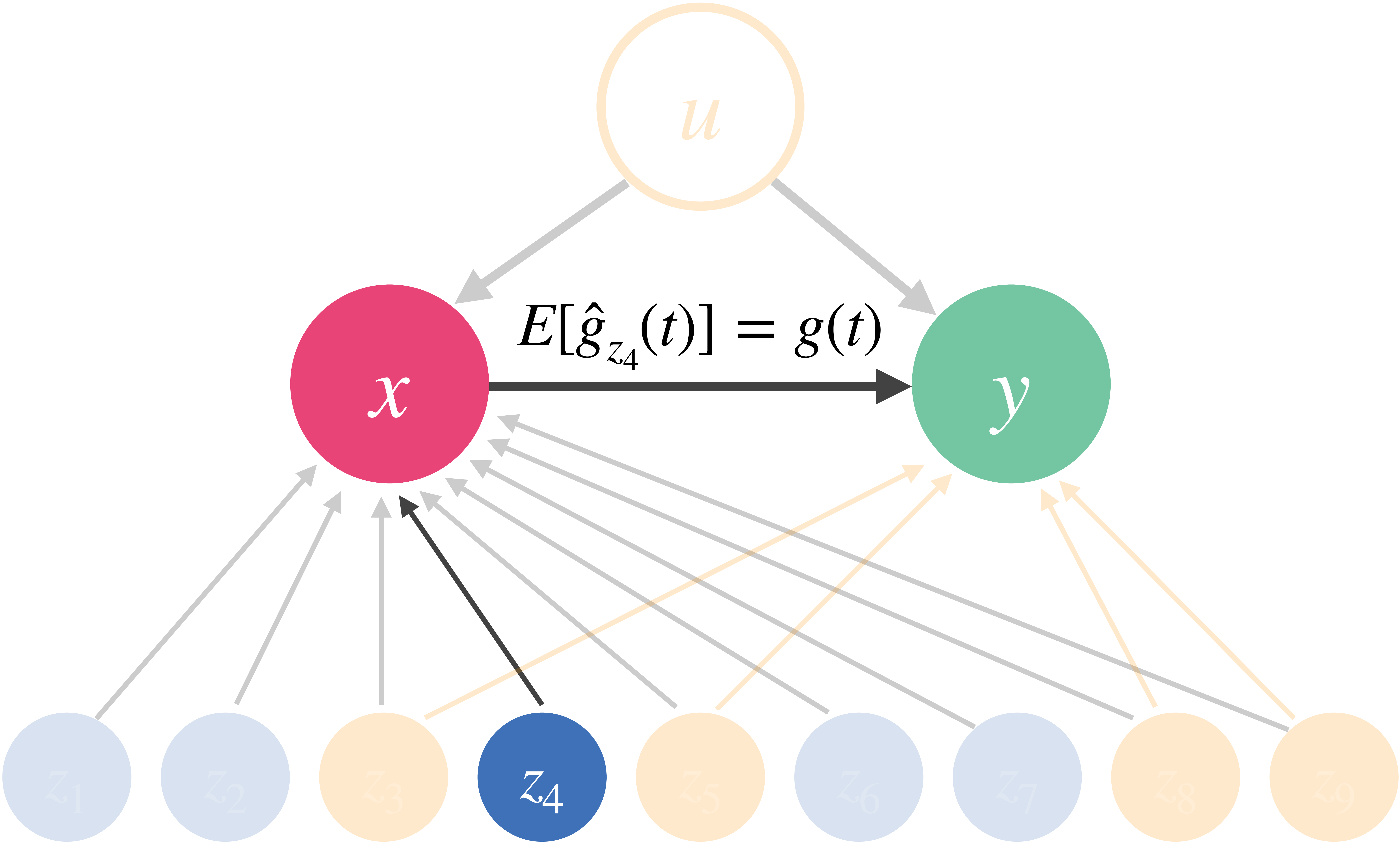
[c.f. Hartwig et al. 2017]

One way to be unbiased



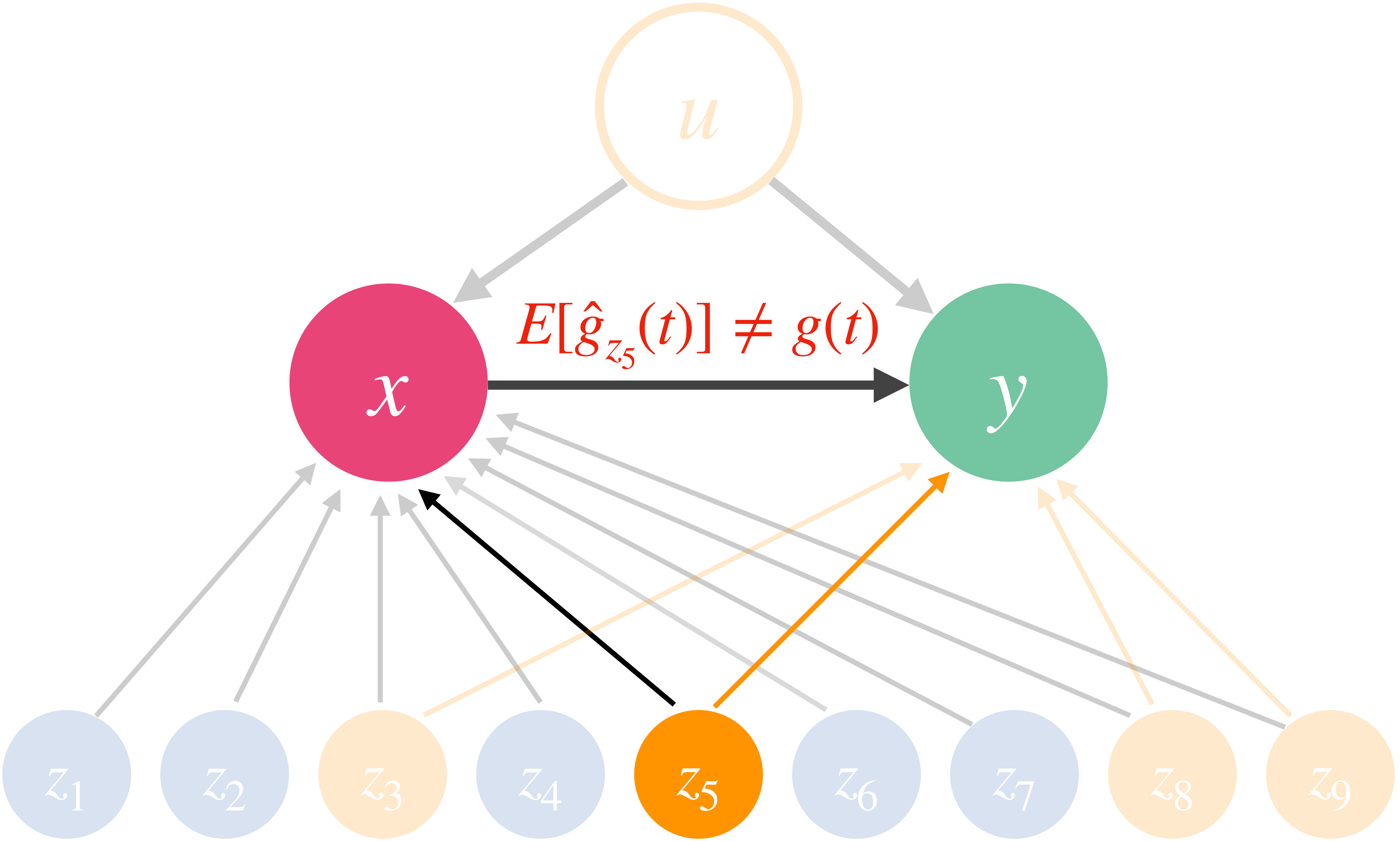
[c.f. Hartwig et al. 2017]

One way to be unbiased



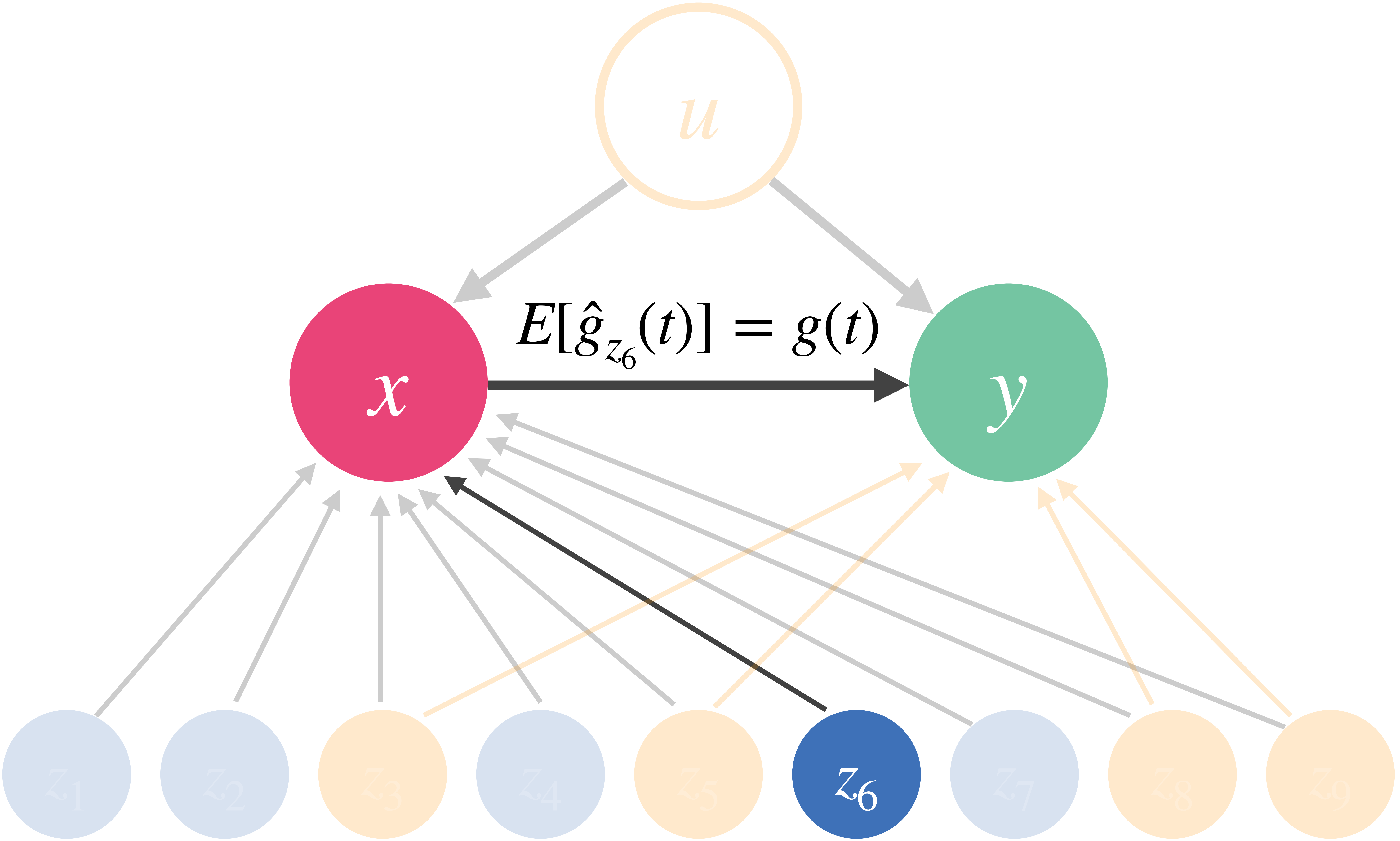
[c.f. Hartwig et al. 2017]

One way to be unbiased



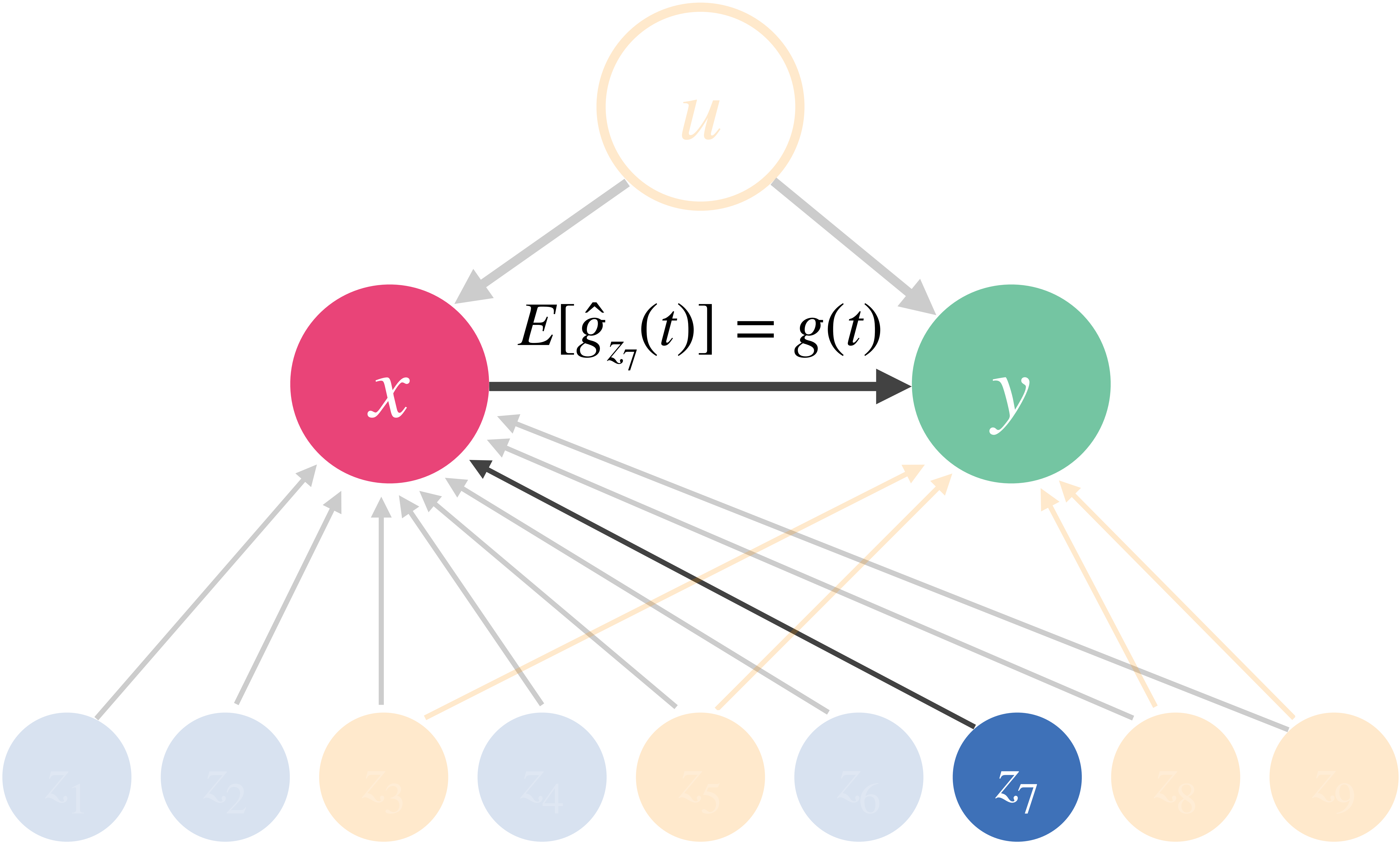
[c.f. Hartwig et al. 2017]

One way to be unbiased



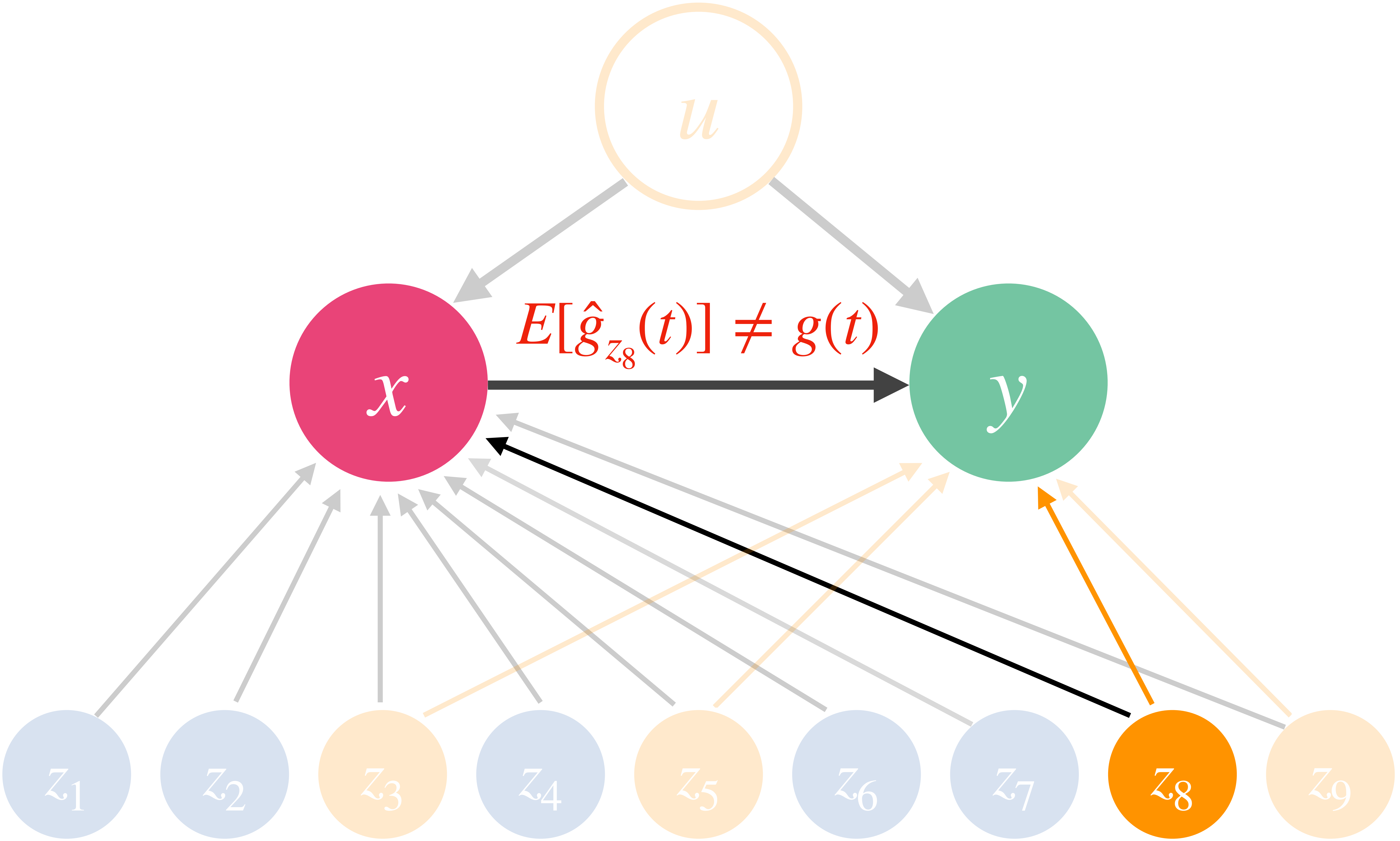
[c.f. Hartwig et al. 2017]

One way to be unbiased



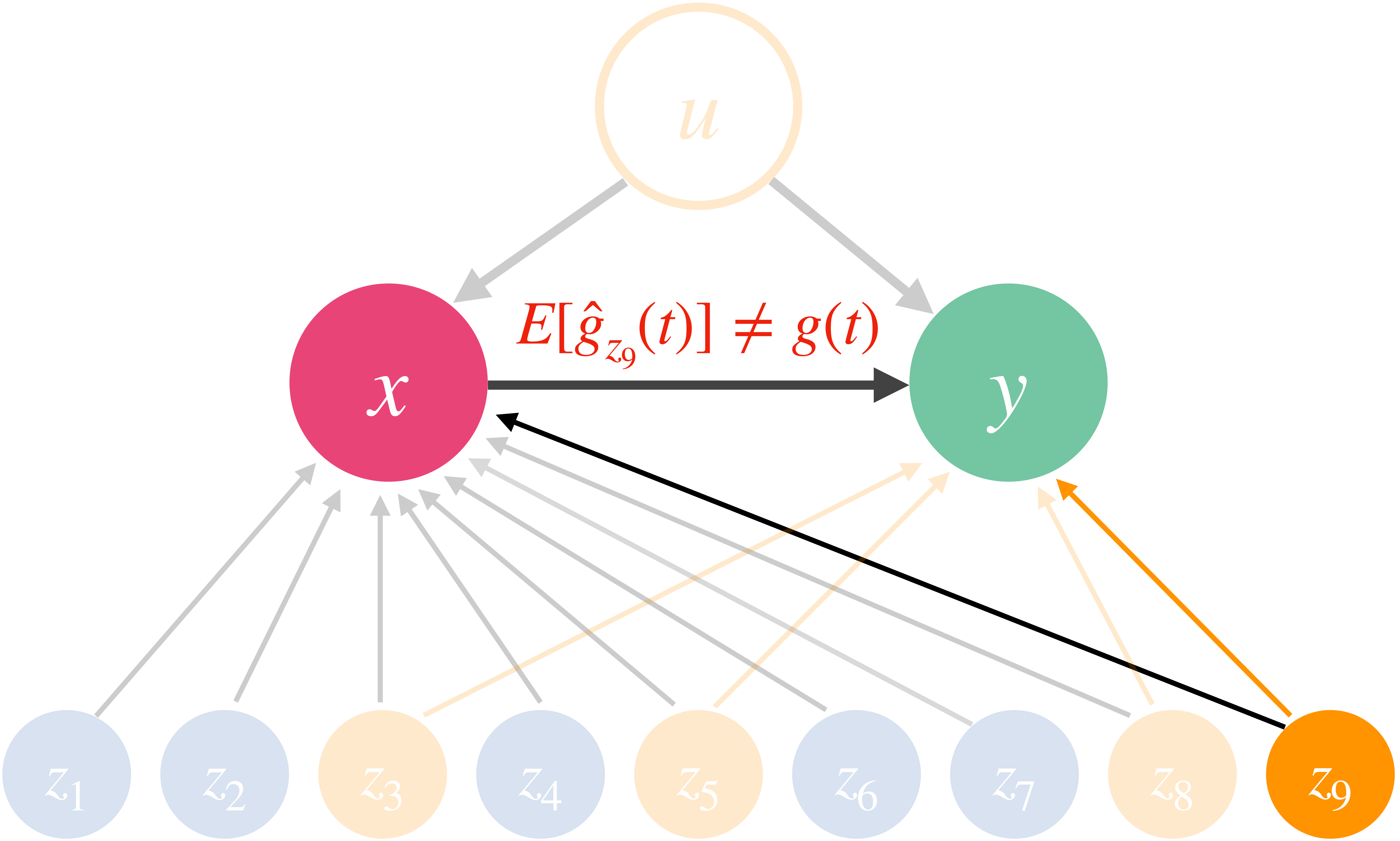
[c.f. Hartwig et al. 2017]

One way to be unbiased



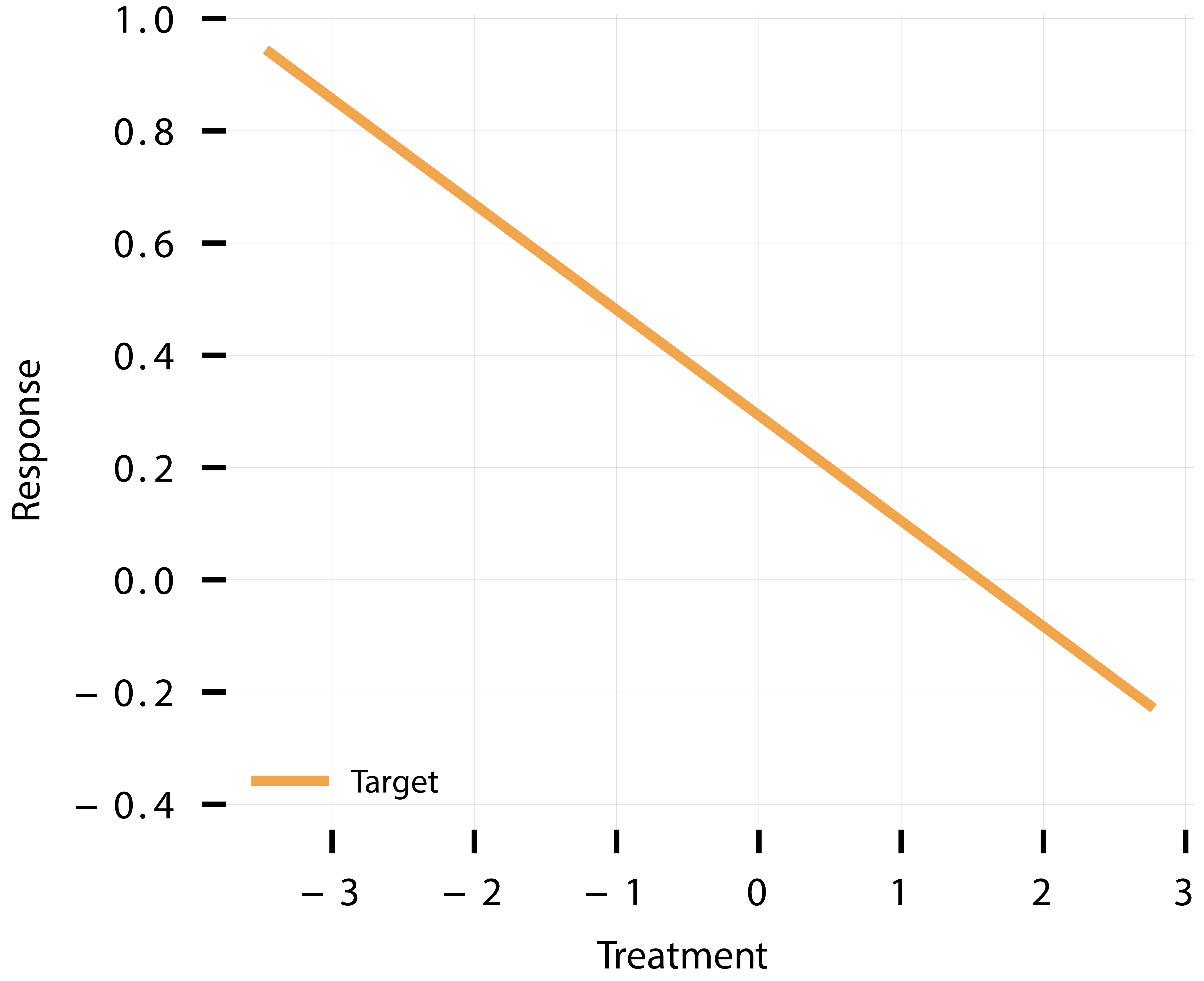
[c.f. Hartwig et al. 2017]

One way to be unbiased



[c.f. Hartwig et al. 2017]

ModelIV



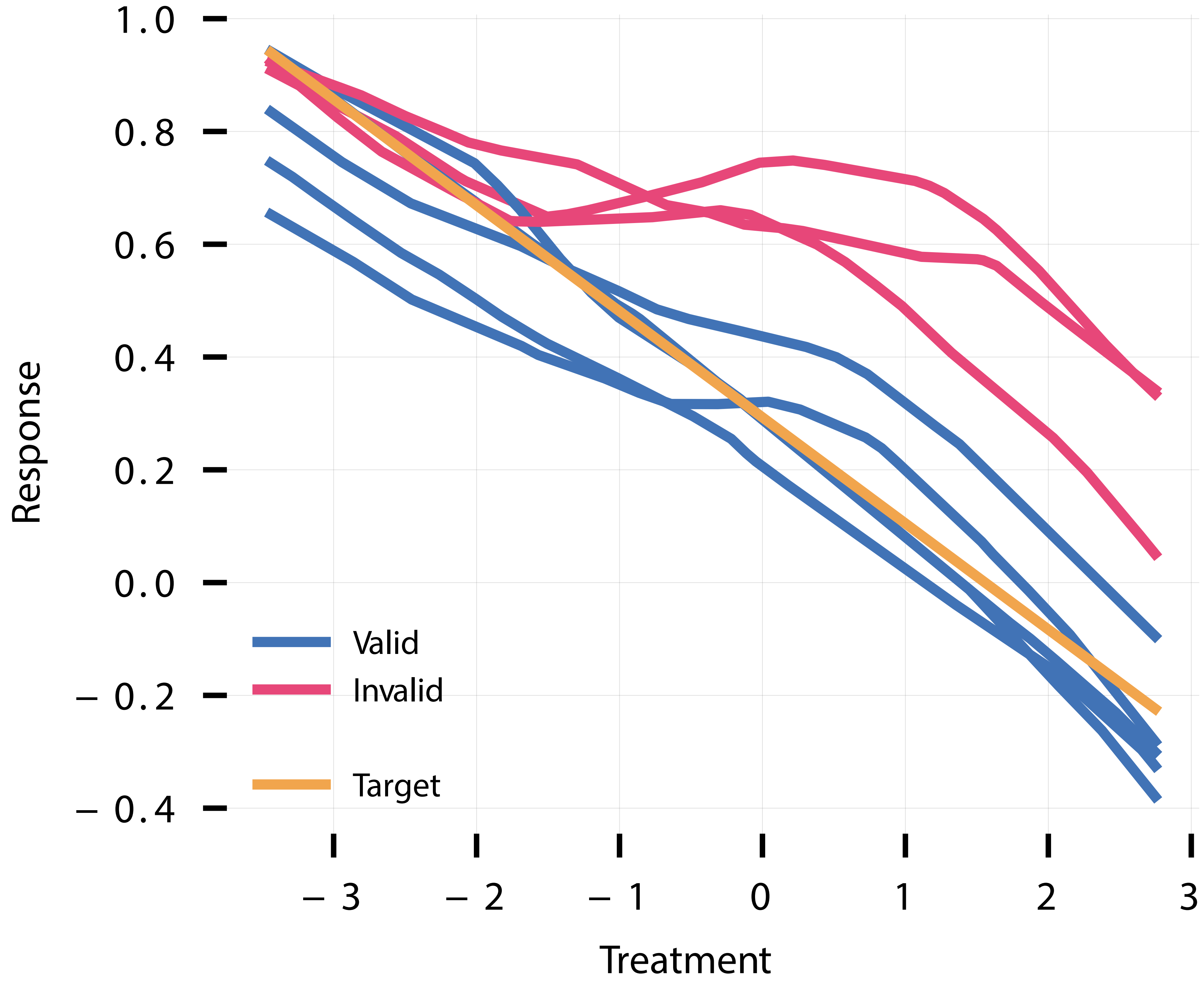
ModelIV algorithm in two steps:

Input: lower bound on the number of valid instruments, V

1. If you have k instrumental variables, fit an **ensemble** of k different instances of DeepIV / DeepGMM / Kernel IV / etc.

2. Output the 'Venter **mode**' of the ensemble (mean of V closest estimates).

ModelIV



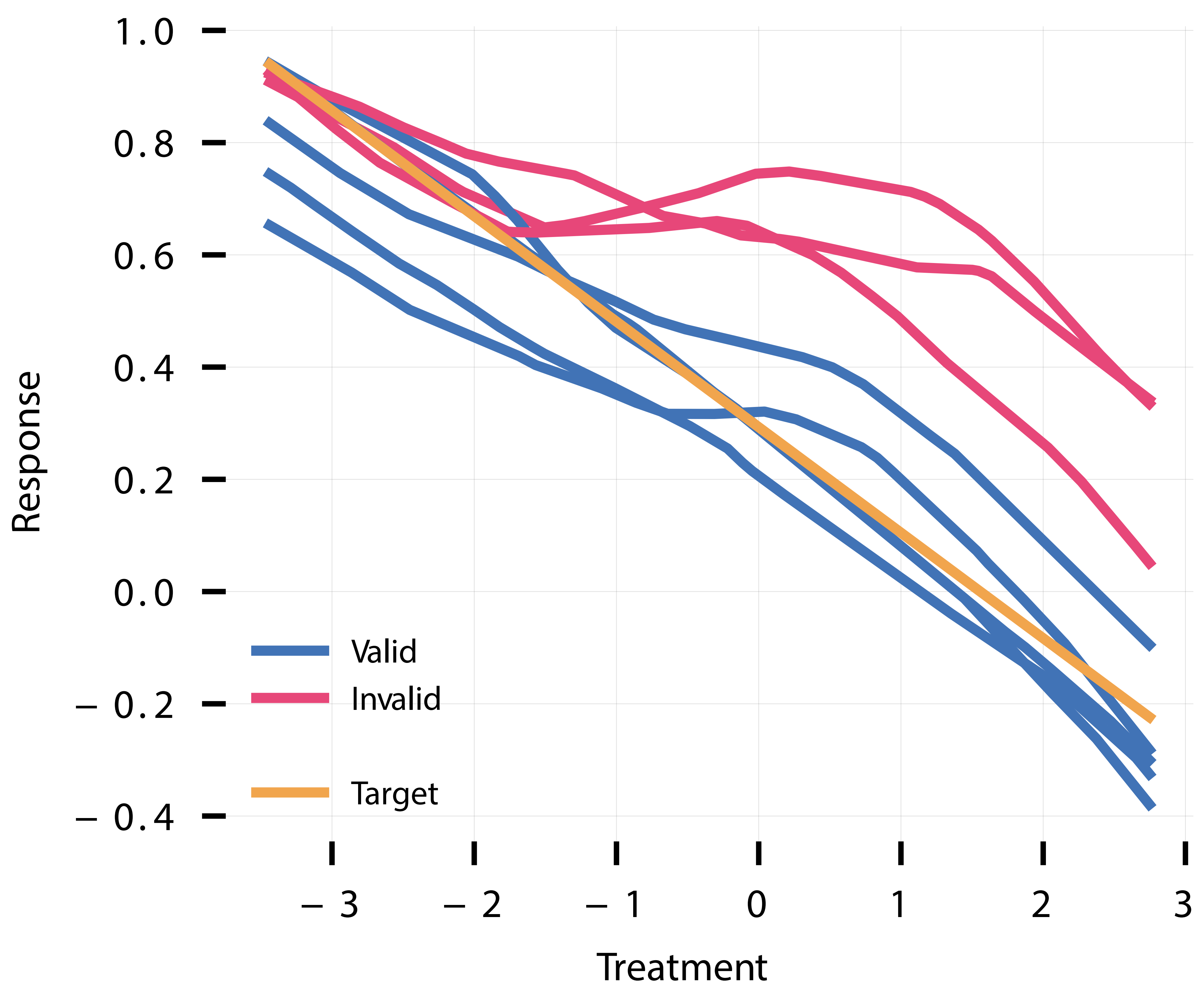
ModelIV algorithm in two steps:

Input: lower bound on the number of valid instruments, V

1. If you have k instrumental variables, fit an **ensemble** of k different instances of DeepIV / DeepGMM / Kernel IV / etc.

2. Output the 'Venter **mode**' of the ensemble (mean of V closest estimates).

ModelIV

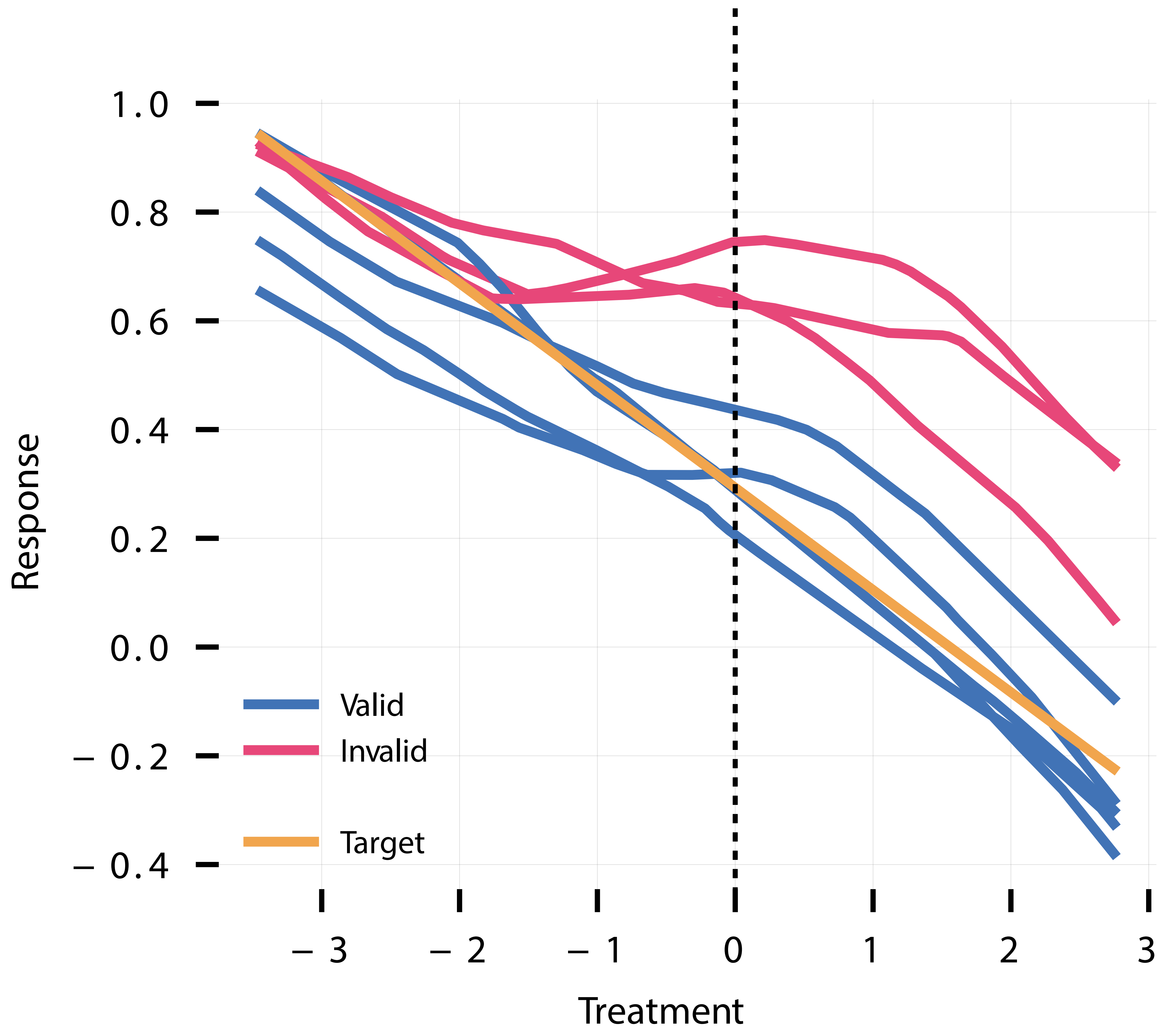


ModelIV algorithm in two steps:

Input: lower bound on the number of valid instruments, V

1. If you have k instrumental variables, fit an **ensemble** of k different instances of DeepIV / DeepGMM / Kernel IV / etc.
2. Output the 'Venter **mode**' of the ensemble (mean of V closest estimates).

ModelIV



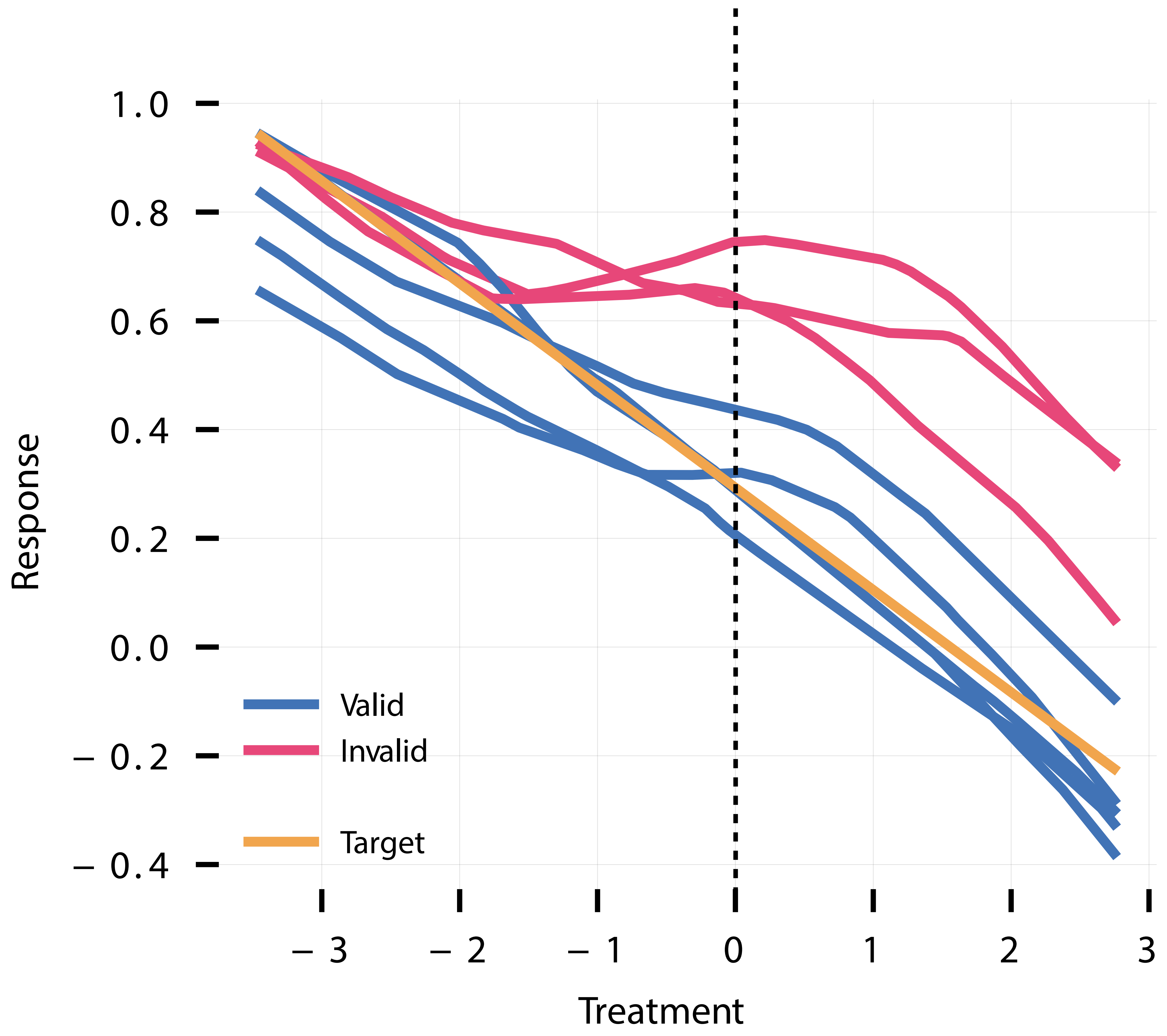
ModelIV algorithm in two steps:

Input: lower bound on the number of valid instruments, V

1. If you have k instrumental variables, fit an **ensemble** of k different instances of DeepIV / DeepGMM / Kernel IV / etc.

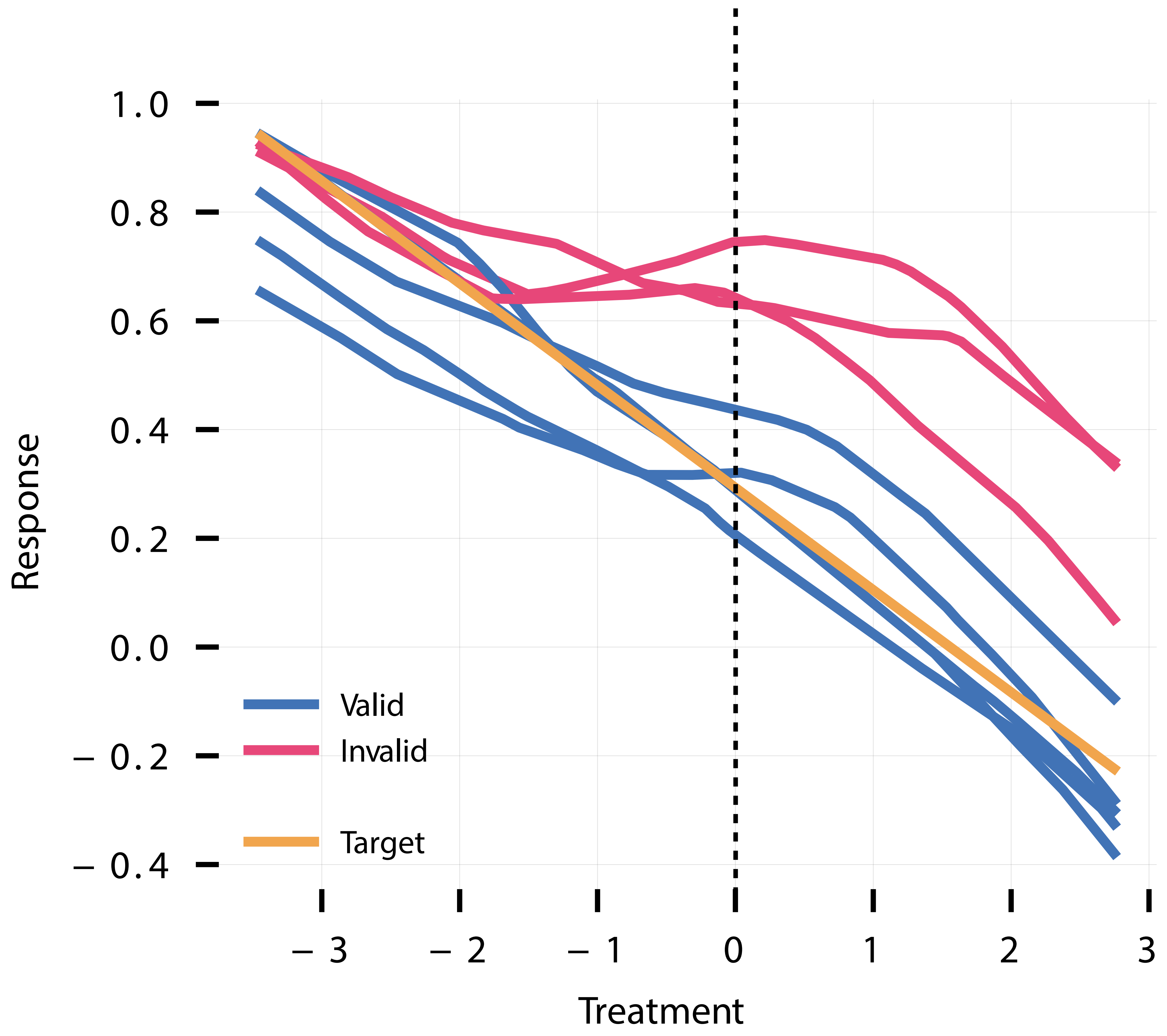
2. Output the 'Venter **mode**' of the ensemble (mean of V closest estimates).

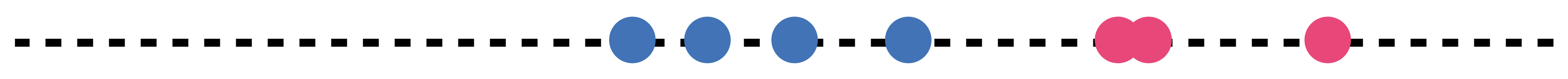
ModelV



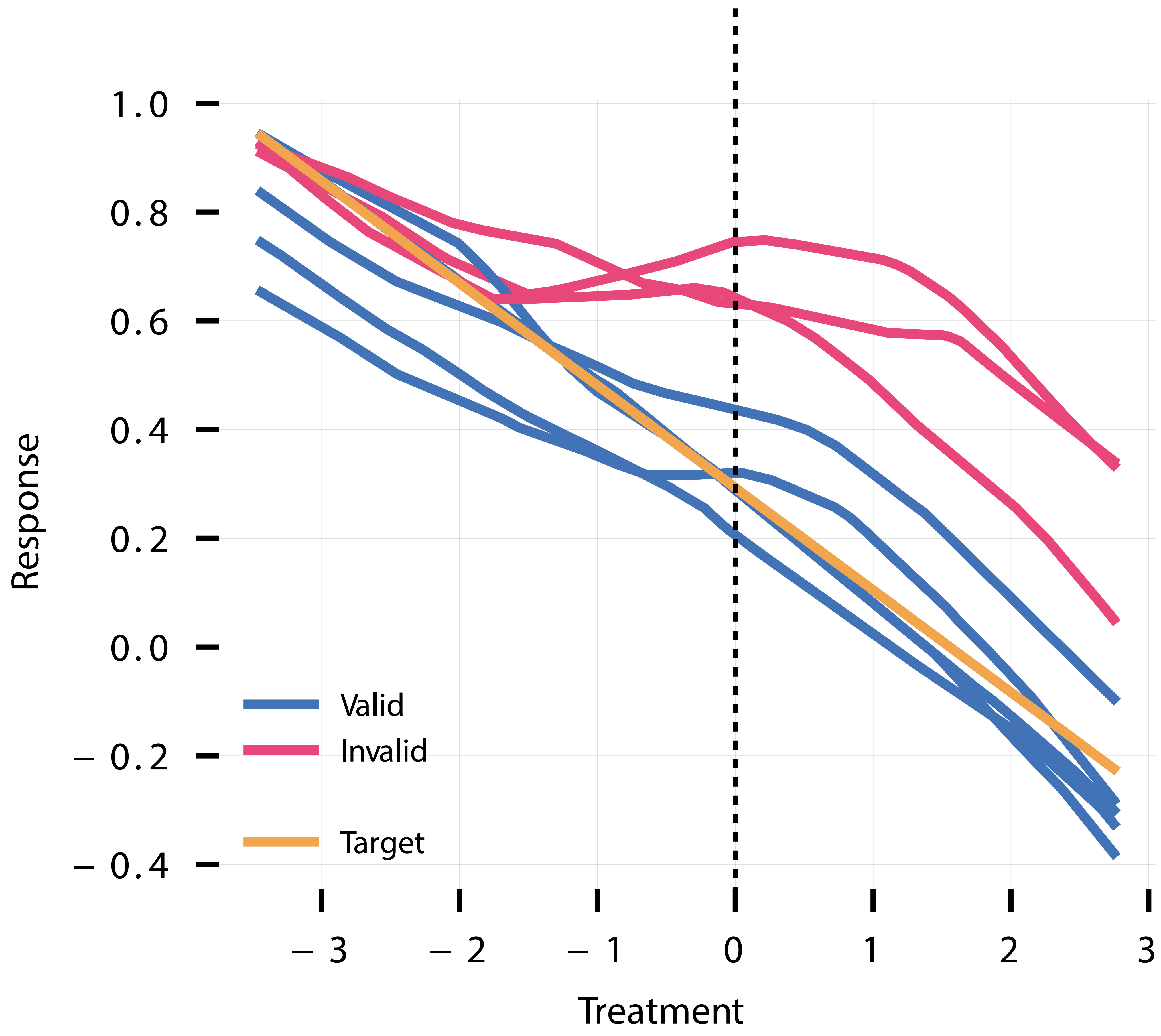
1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
Valid
Invalid
Target
– 3
– 2
– 1
0
1
2
3
Treatment

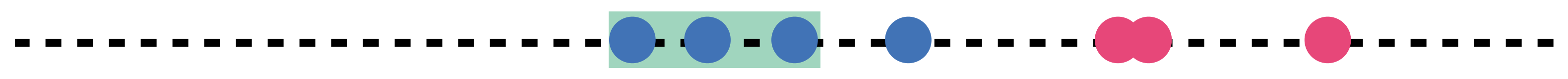
ModelV


1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
– 3
– 2
– 1
0
1
2
3
Treatment
Valid
Invalid
Target

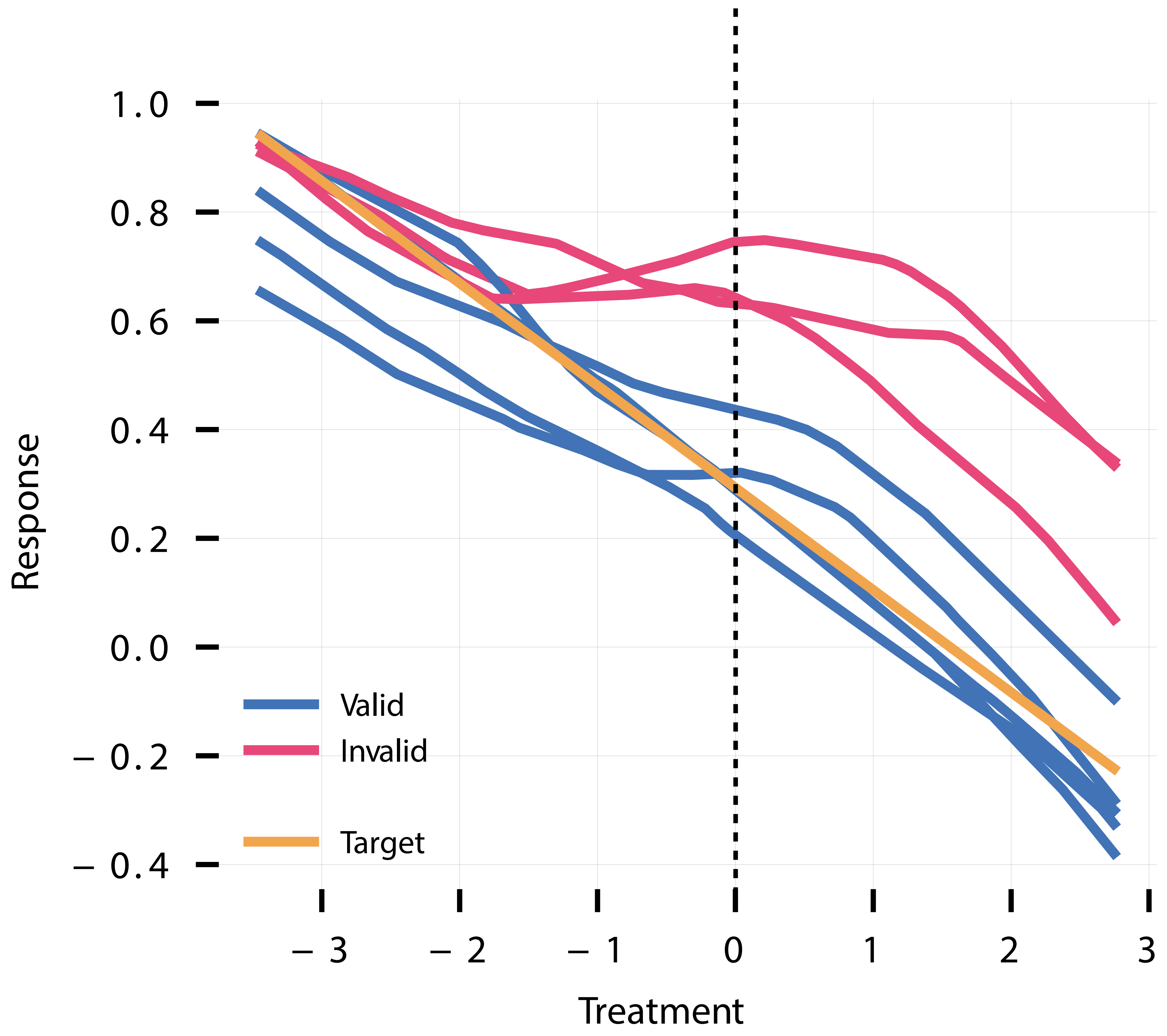


ModelV

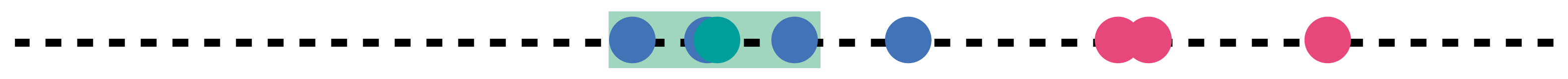

1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
Valid
Invalid
Target
– 3
– 2
– 1
0
1
2
3
Treatment



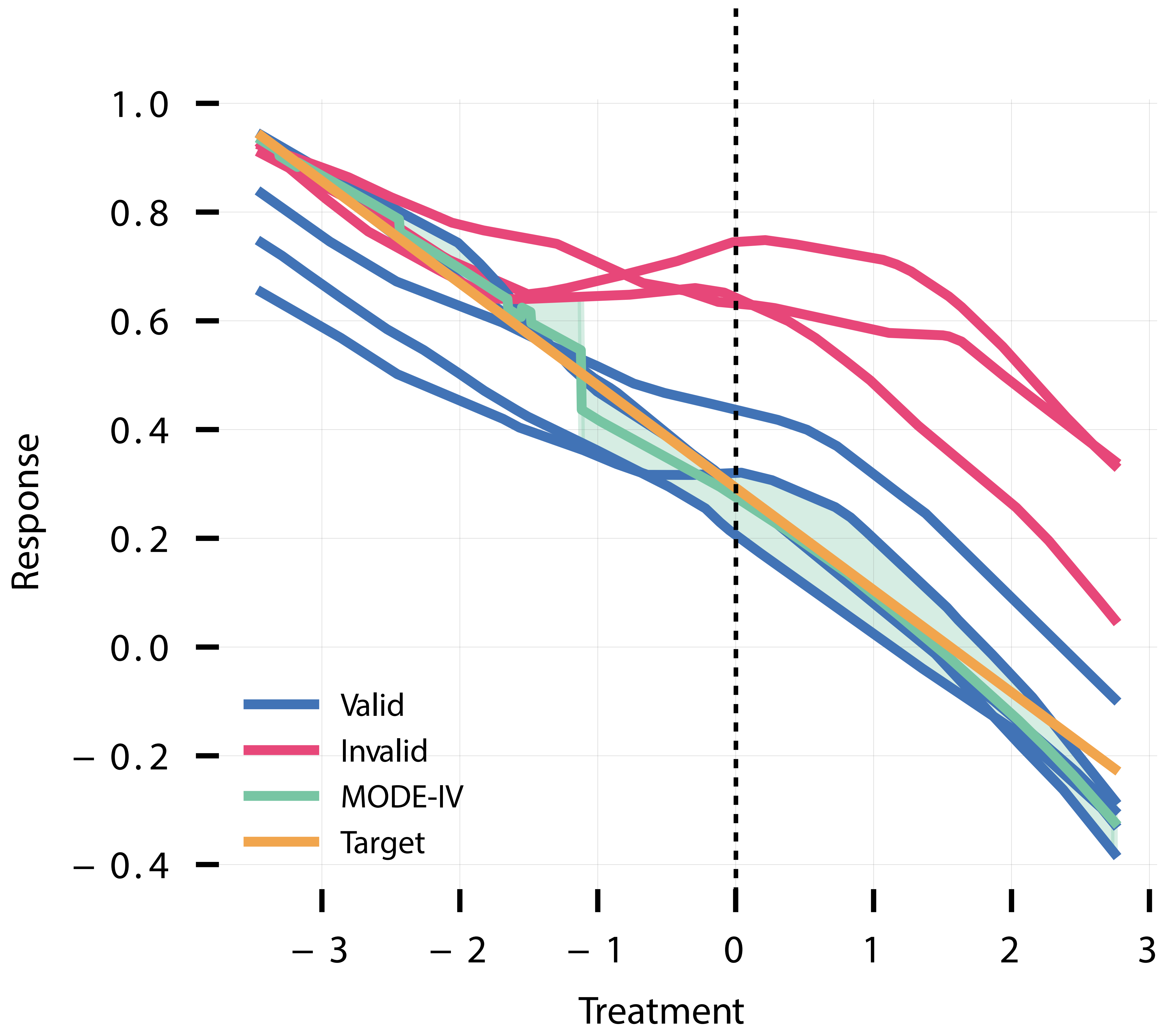
ModelV



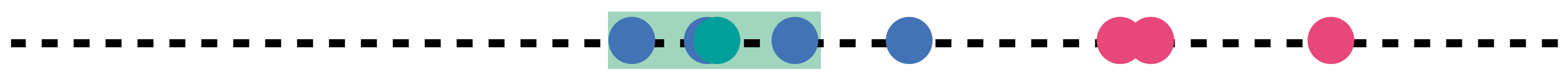
1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
– 3
– 2
– 1
0
1
2
3
Treatment
Valid
Invalid
Target



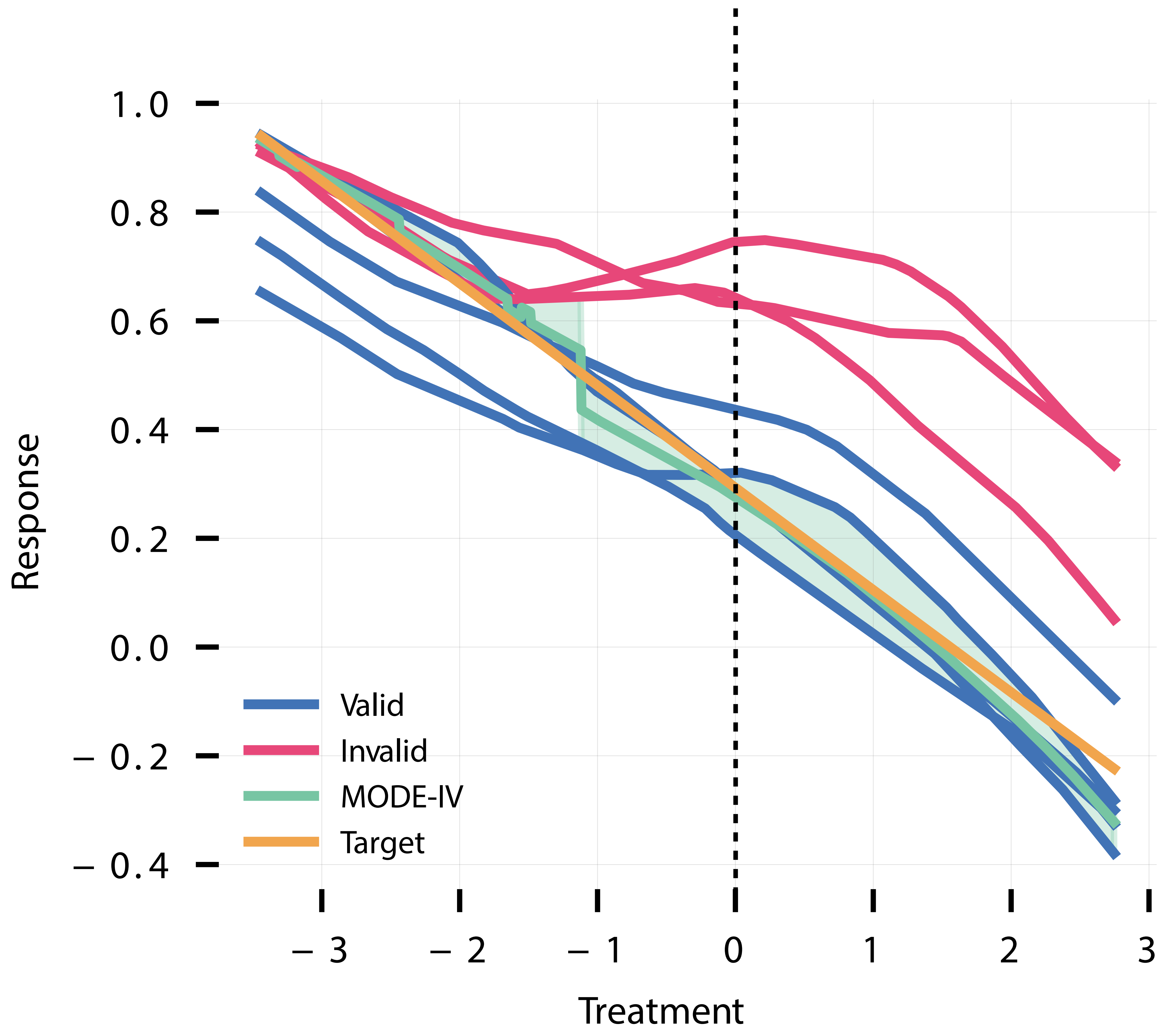
ModelV



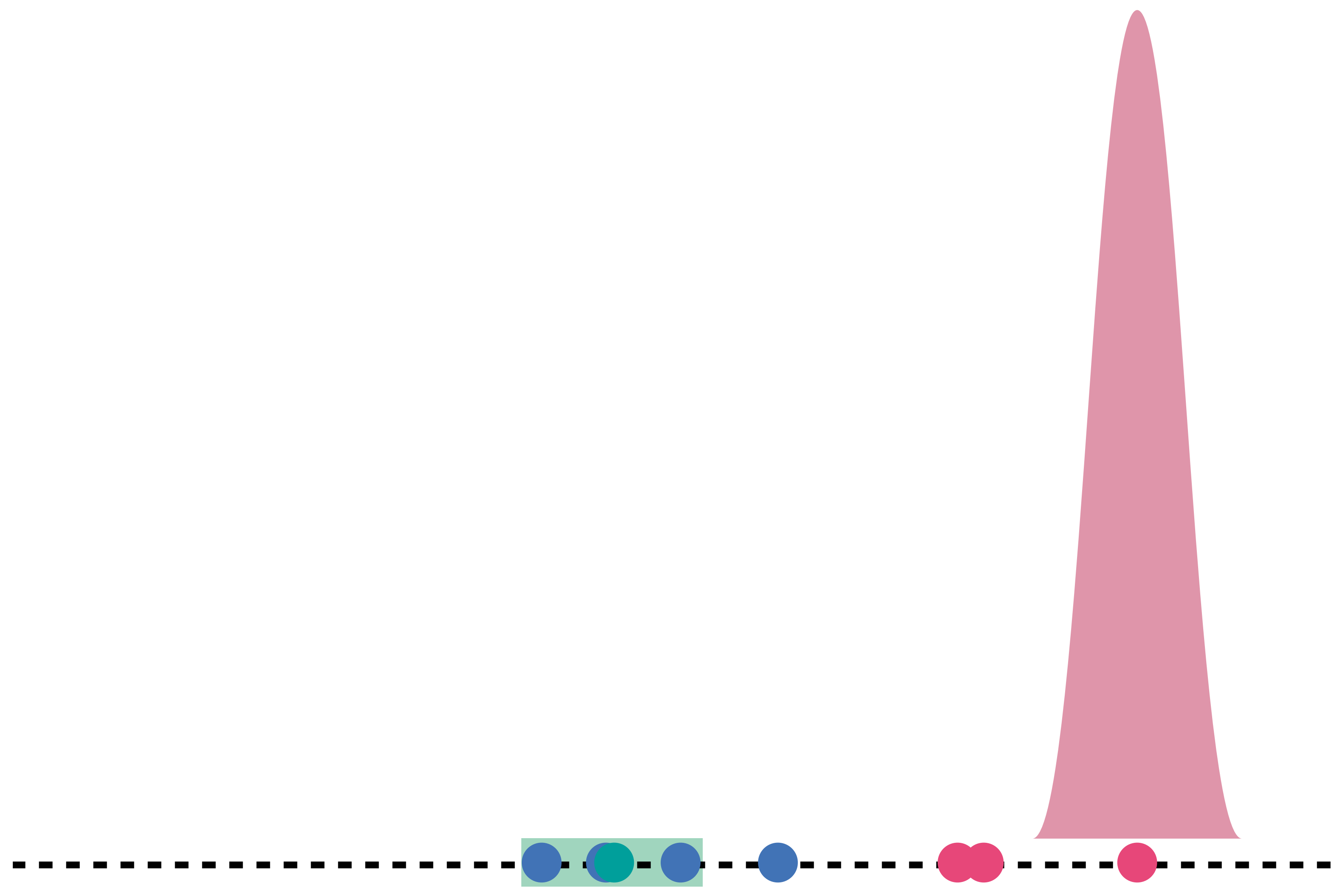
1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
Valid
Invalid
MODE-IV
Target
– 3
– 2
– 1
0
1
2
3
Treatment



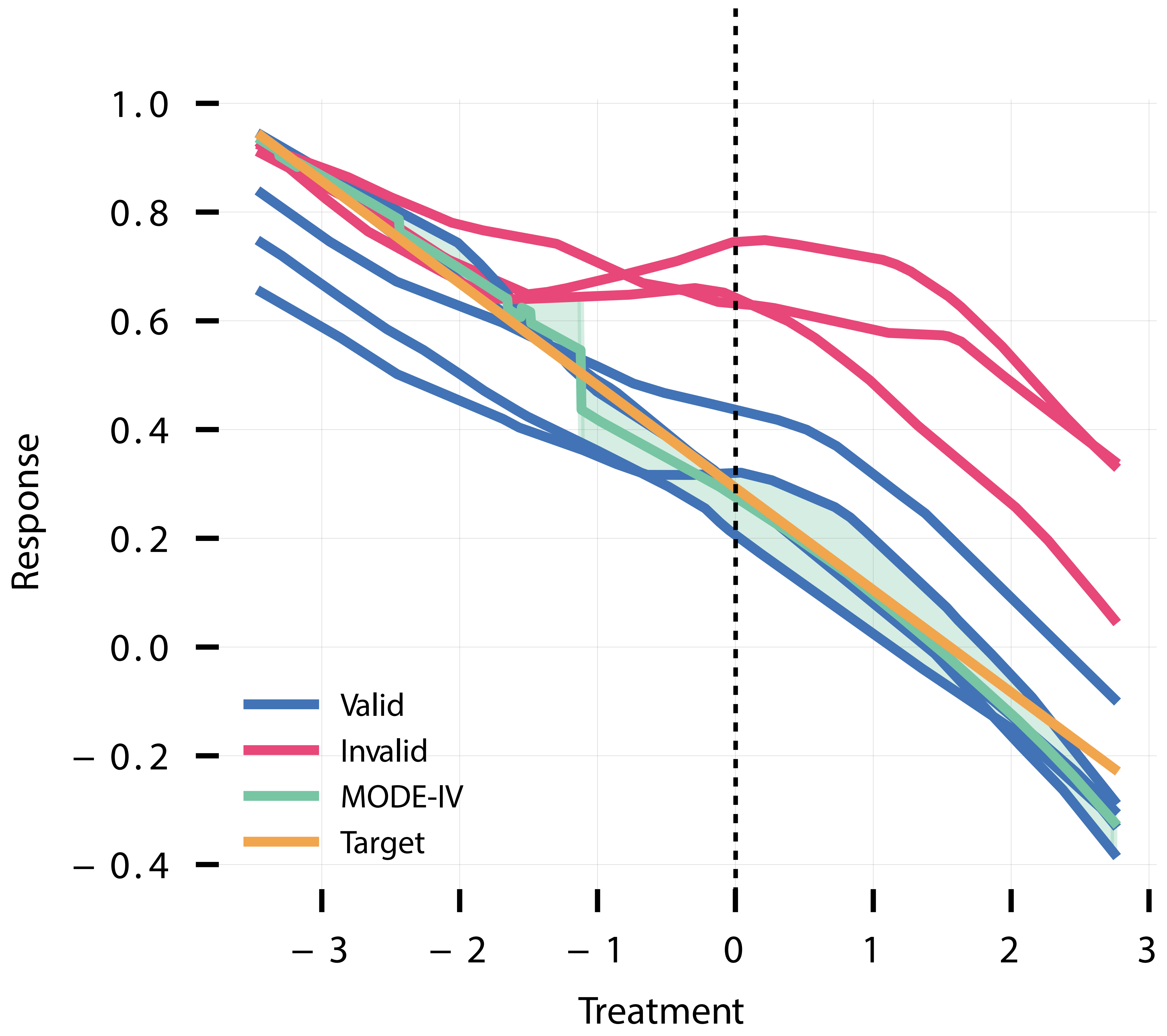
ModelV



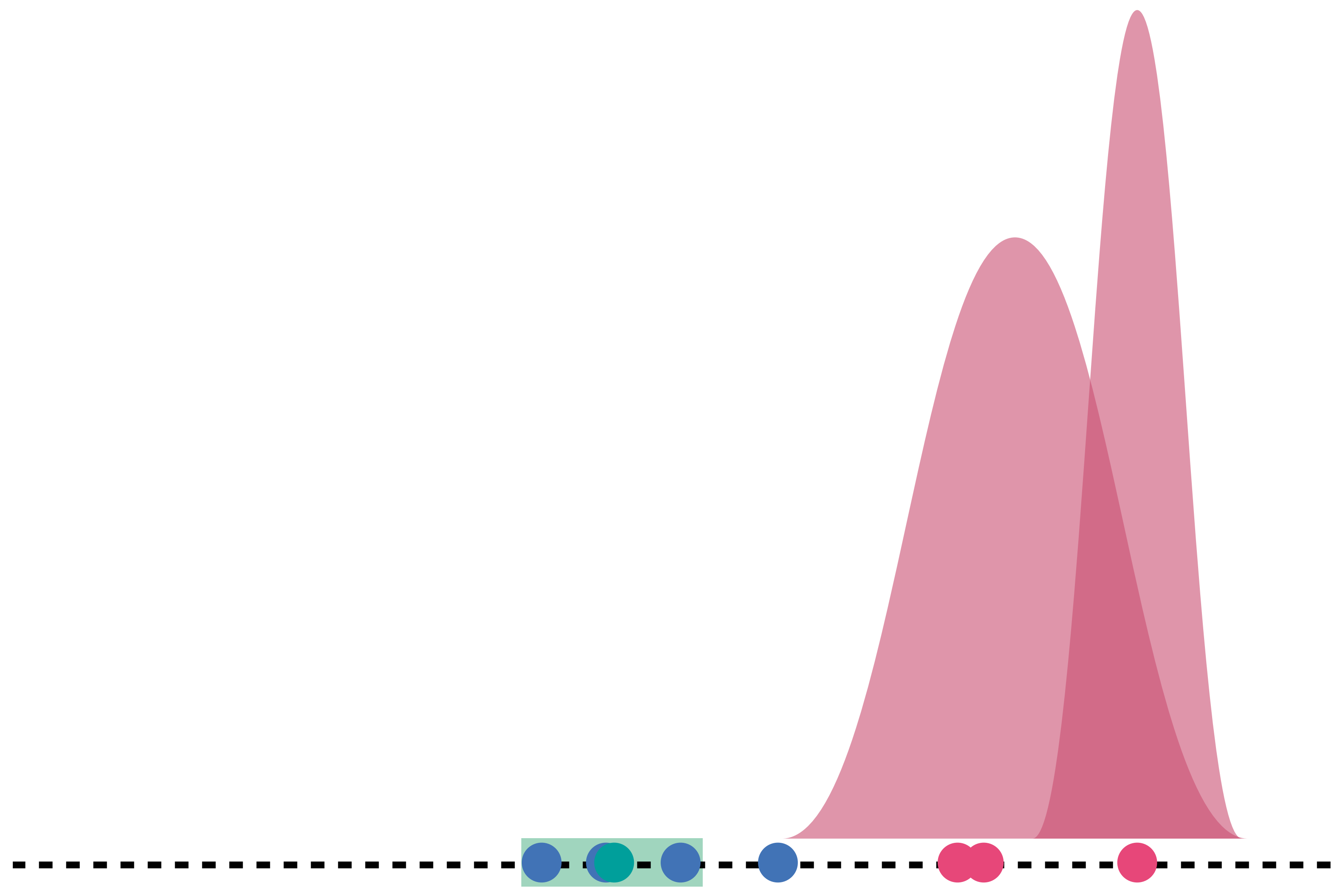
1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
– 3
– 2
– 1
0
1
2
3
Treatment
Valid
Invalid
MODE-IV
Target



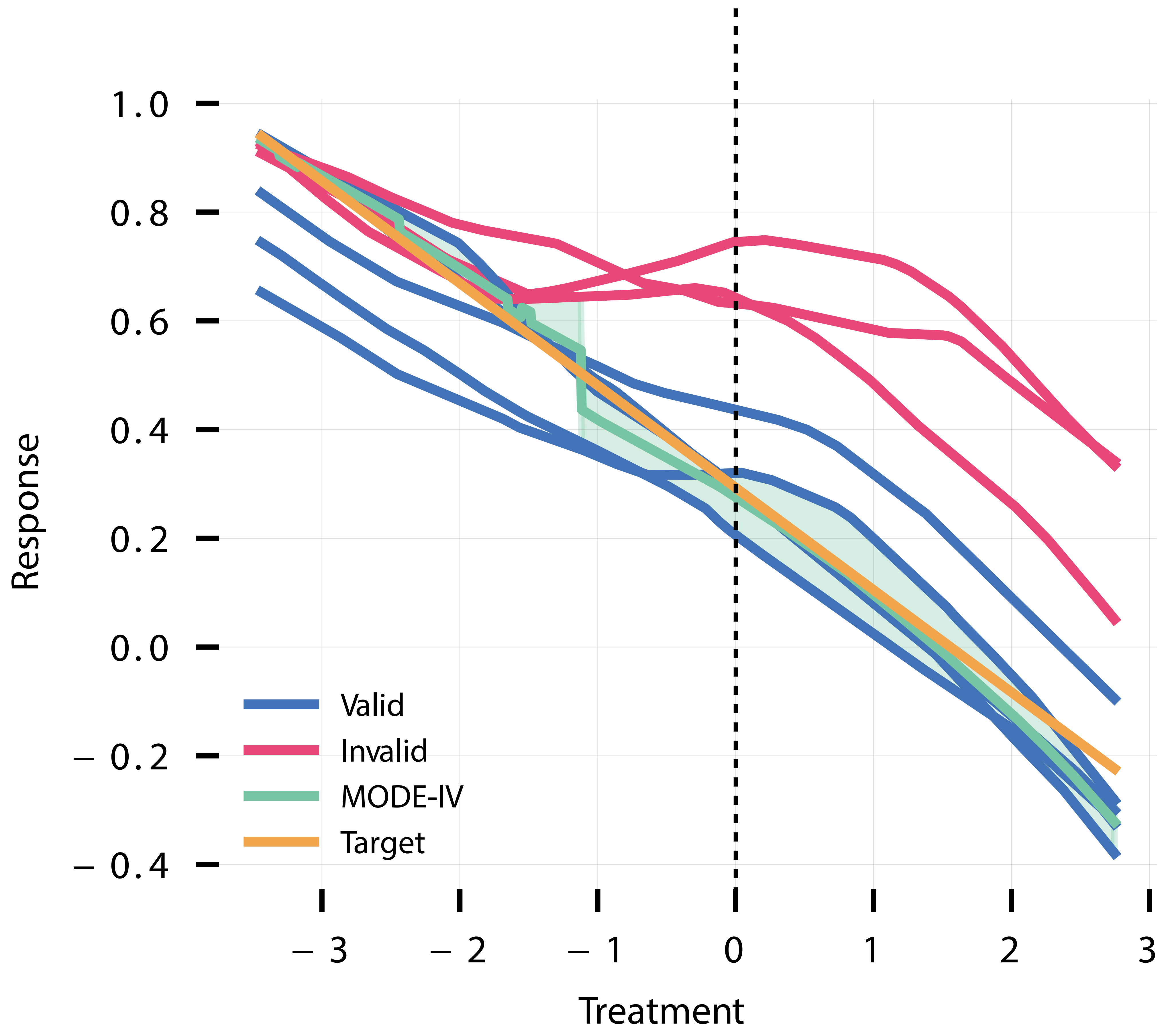
ModelV



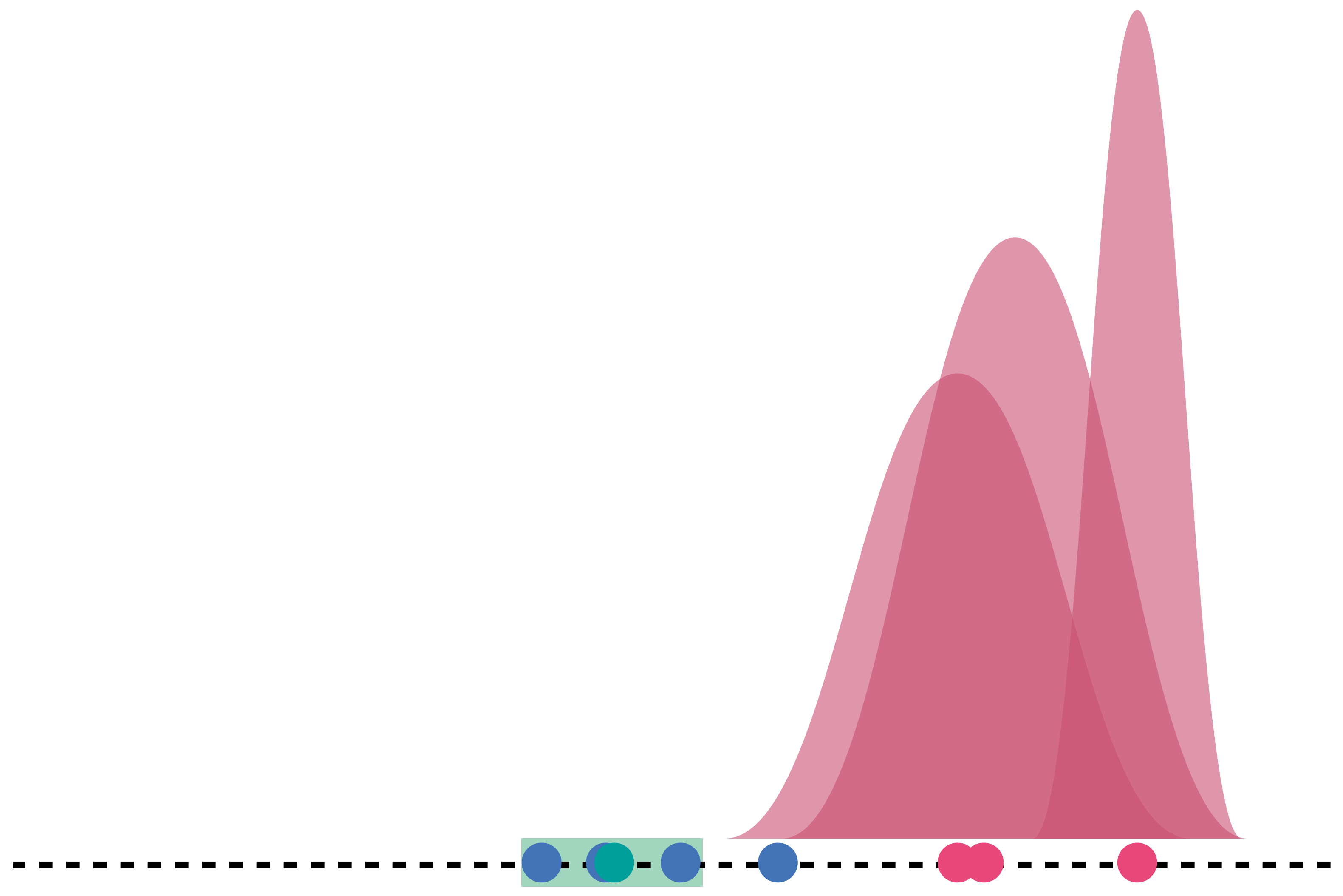
1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
– 3
– 2
– 1
0
1
2
3
Treatment
Valid
Invalid
MODE-IV
Target



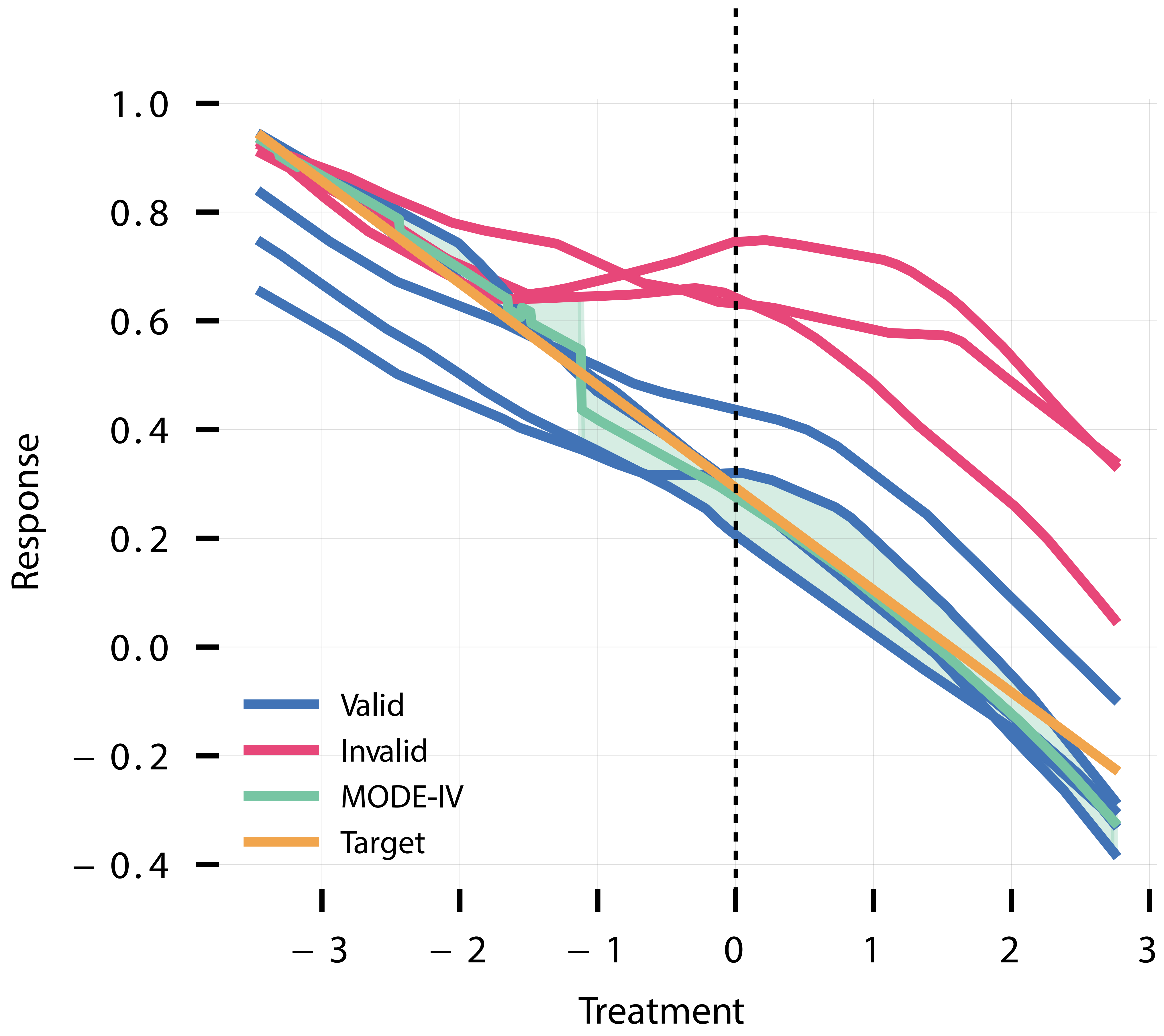
ModelV



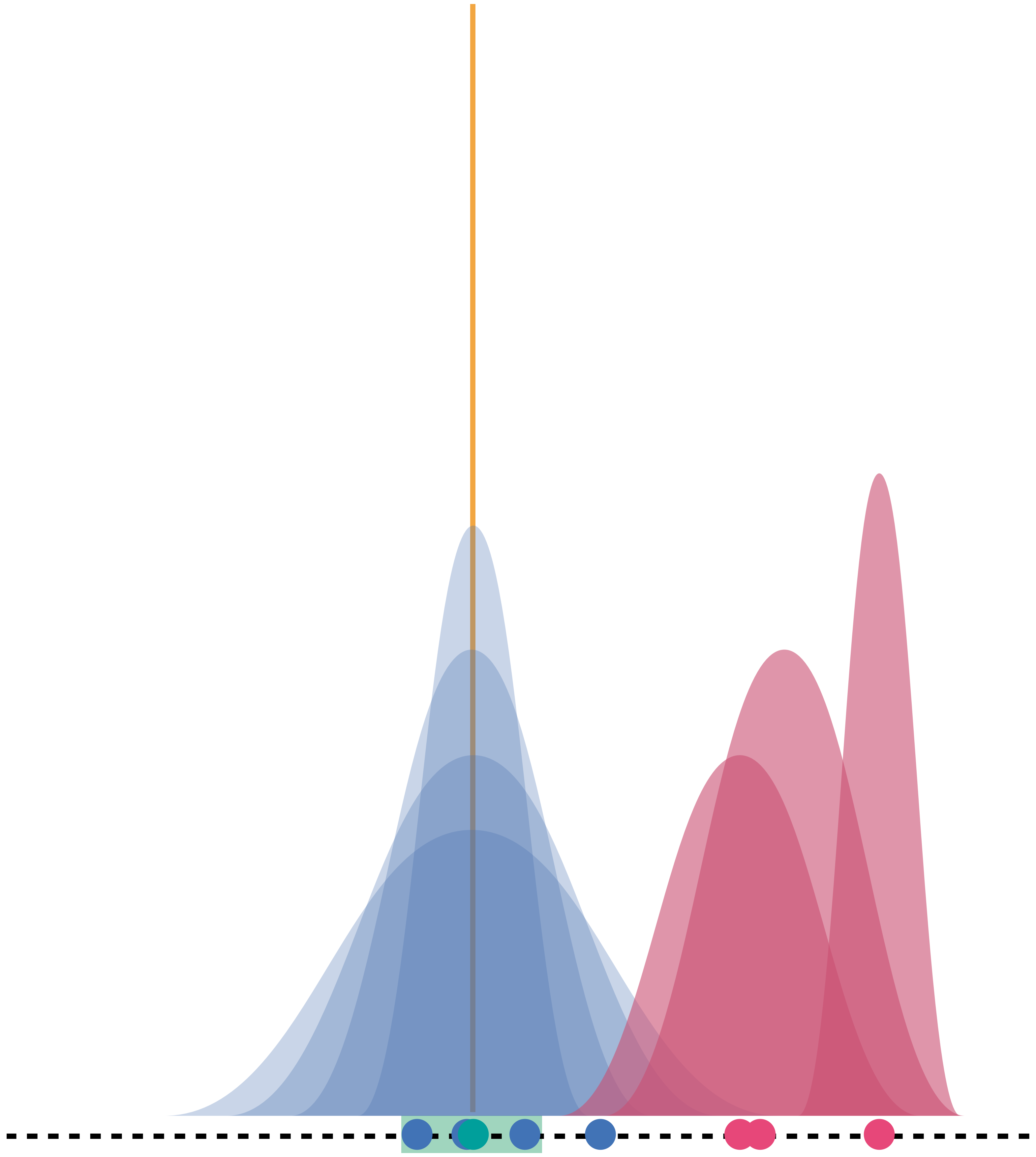
Response
1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Valid
Invalid
MODE-IV
Target
– 3
– 2
– 1
0
1
2
3
Treatment



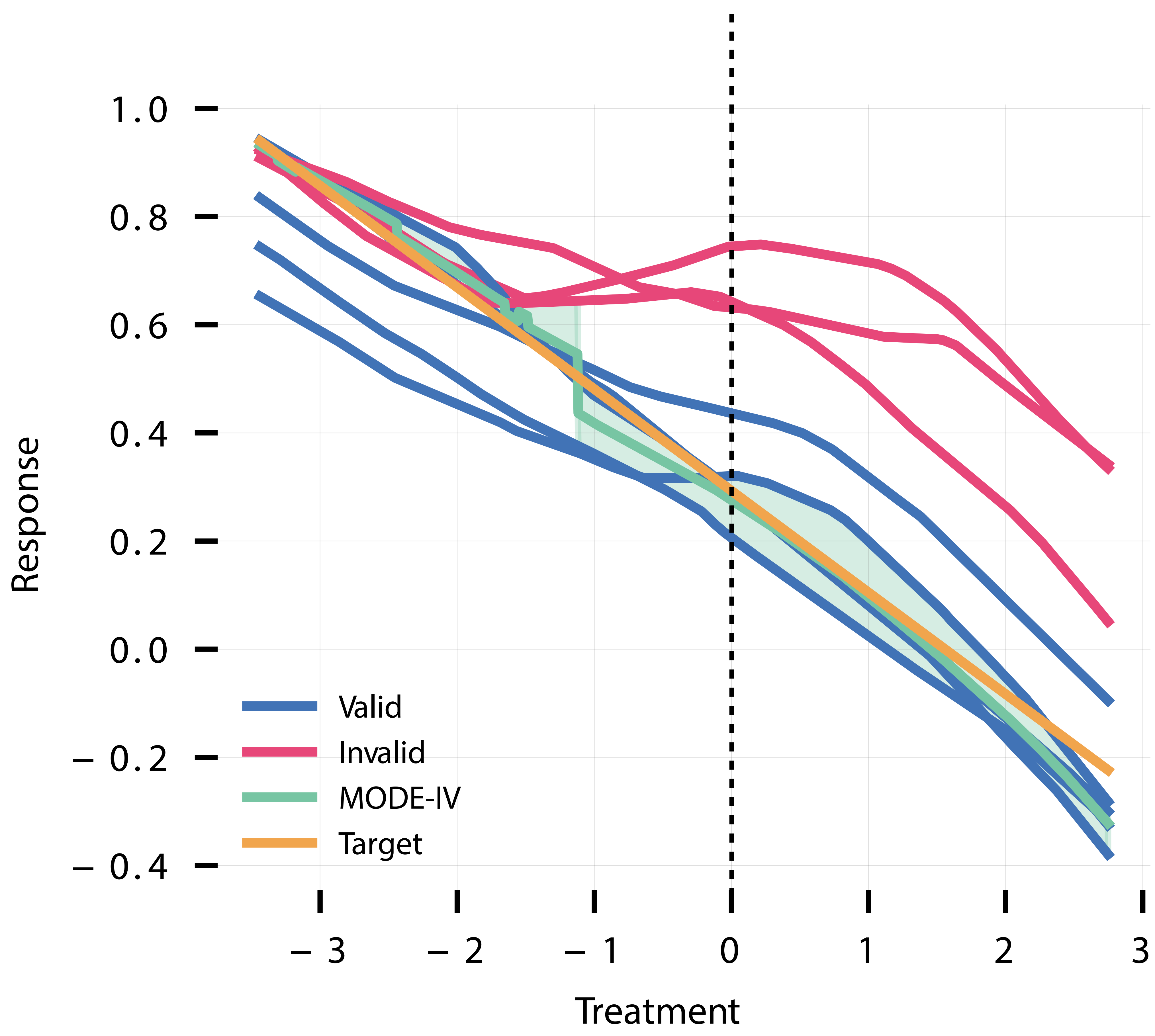
ModelV

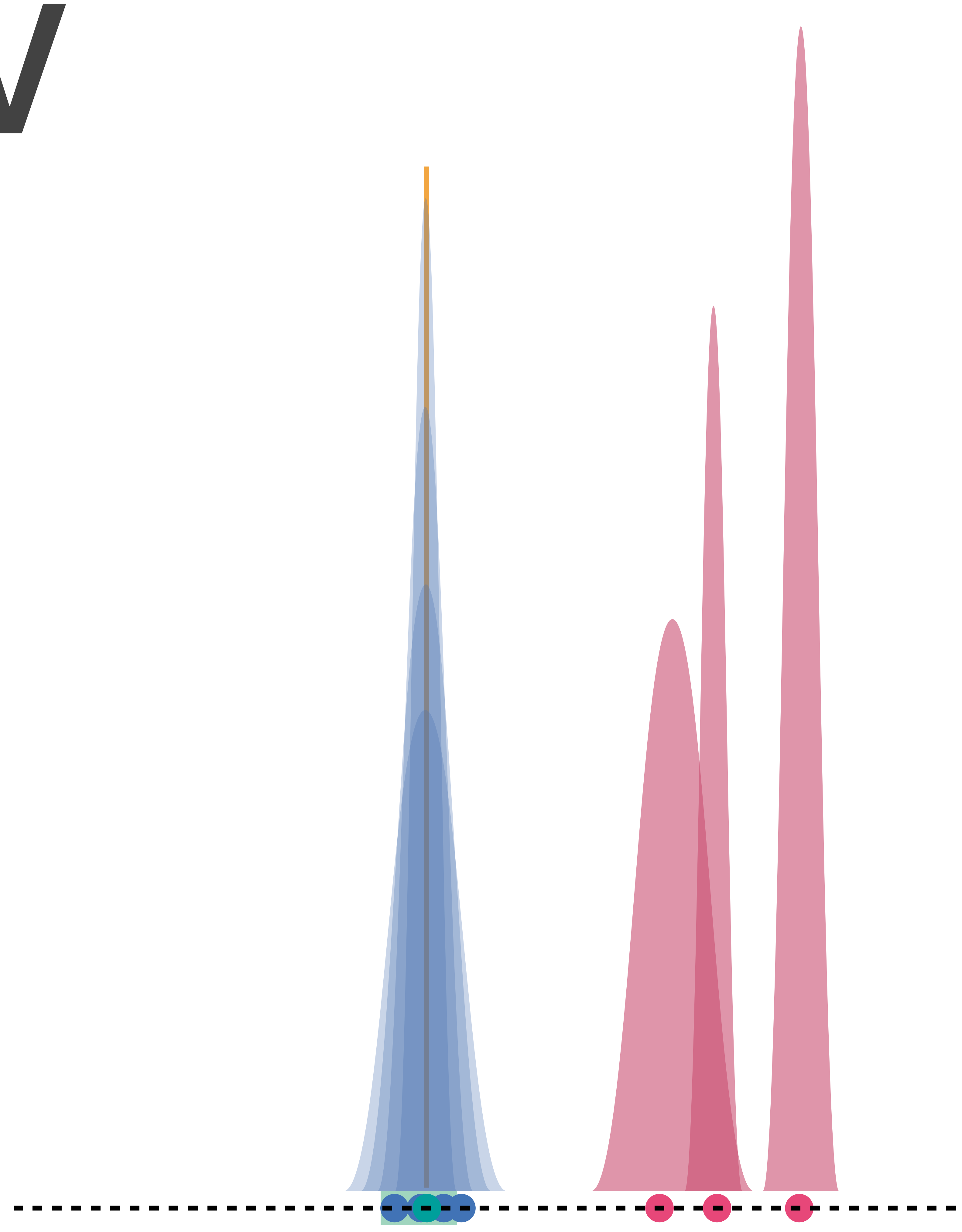


1.0
0.8
0.6
0.4
0.2
0.0
– 0.2
– 0.4
Response
– 3
– 2
– 1
0
1
2
3
Treatment
Valid
Invalid
MODE-IV
Target



ModeIV





***Theorem**: If each estimator is consistent and modal validity hold, ModeIV is a consistent estimator for the true effect $E[y \mid do(p), x]$*

Results summary

ModelV…

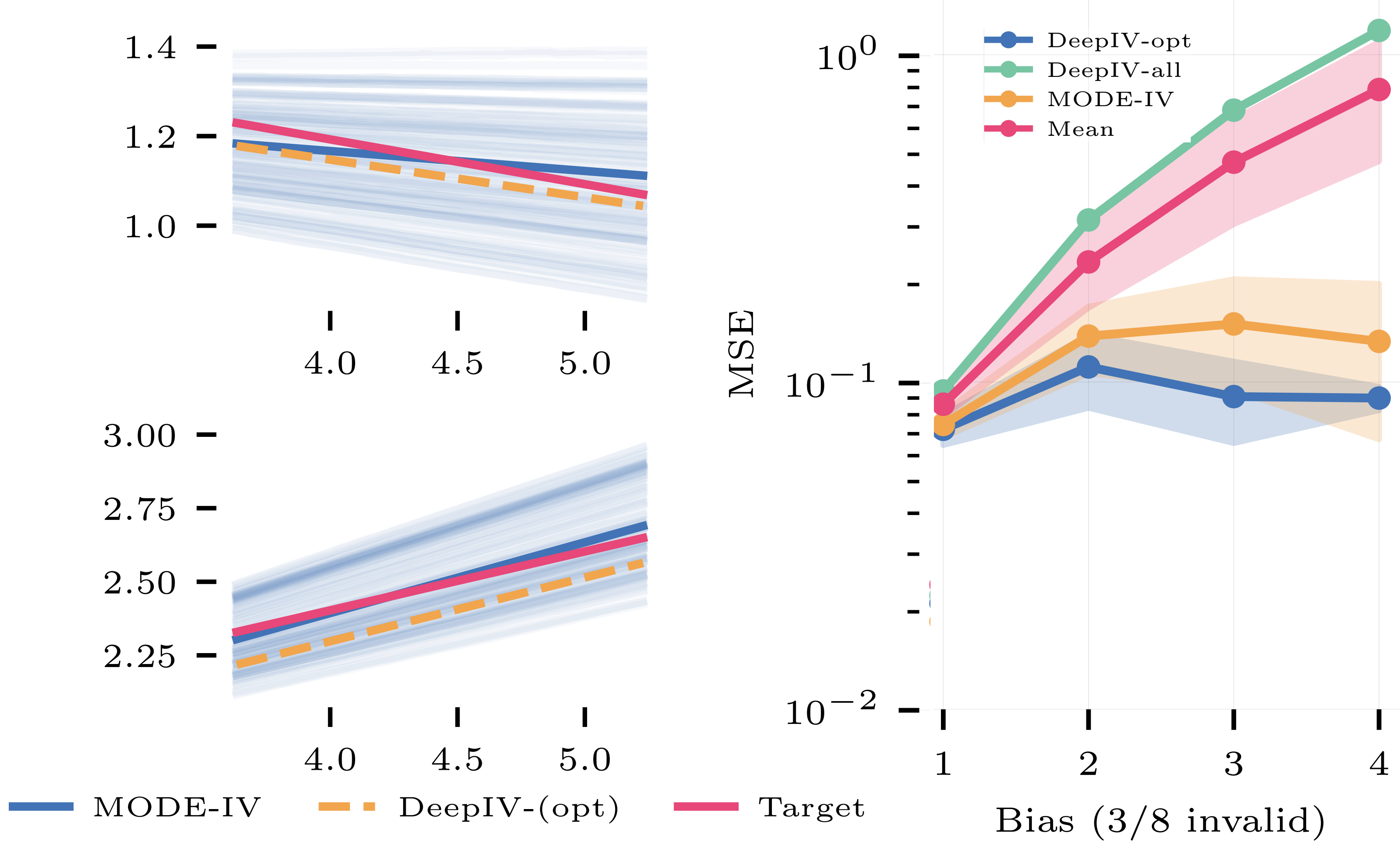
- Performs well in finite sample simulations. Successfully removes most of the bias introduced by invalid instruments.
- Converges at the same rate as the underlying estimators, even on worst case distributions.

Theorem 2. *For some test point* (t, x), *let* $\mathcal{Z} = \{\hat{\beta}_1, \ldots, \hat{\beta}_k\}$ *be* k *estimates of the causal effect of* t *at* x. *Assume,*

*[**Bounded estimates**] Each estimate is bounded by some constants,* $[a_i, b_i]$

*[**Convergent estimators**] Each estimator converges in mean squared error at a rate* n^{-r} *(where* $r = \frac{1}{2}$ *if the estimator achieves the parametric rate), and hence each estimator has finite variance, Var*$(\hat{\beta}_i) = \frac{\sigma_i}{n^{-2r}}$ *for some* σ_i.

Then, if $\sigma = \max_{i \in \mathcal{V}} \sigma_i$ *there exists a,* C*, such that* $E[(\text{ModeIV}(\mathcal{Z}) - \beta)^2 - (\frac{1}{v} \sum_{i \in \mathcal{V}} \hat{\beta}_i - \beta)^2] \leq 9kC\sigma n^{-r}$.



Summary

- Instrumental variable approaches allow you to estimate causal effects with unobserved confounding.
- ModelV is the first nonparametric procedure that is robust to invalid instruments & is a simple black box procedure.