Bayesian regression modeling: Theory & practice

Part 7: Causal inference & regression modeling

Michael Franke

Causal inference

motivation

- we know: **correlation does not mean causation**
- we want: actionable conclusions
- we use: randomized control trials
- but: **what if we can only observe passively?**

how to disentangle

Simpson’s paradox

Simpson's paradox: Case 1

Gender as a confounder

- 700 patients w/ choice: take drug or not
- variable of interest: recovery rate
- also observed: **gender**

	Drug	No drug
Men	81 / 87 (93%)	234 / 270 (87%)
Women	192 / 263 (73%)	55 / 80 (68%)
Σ	273 / 350 (78%)	289 / 350 (83%)

Case 2

Blood-pressure as a mediator

- same as case 1, but no gender info
- also observed:
  post-treatment blood pressure

	Drug	No drug
Low BP	81 / 87 (93%)	234 / 270 (87%)
High BP	192 / 263 (73%)	55 / 80 (68%)
Σ	273 / 350 (78%)	289 / 350 (83%)

Would you recommend using the drug
in Case 1 and/ or Case 2?

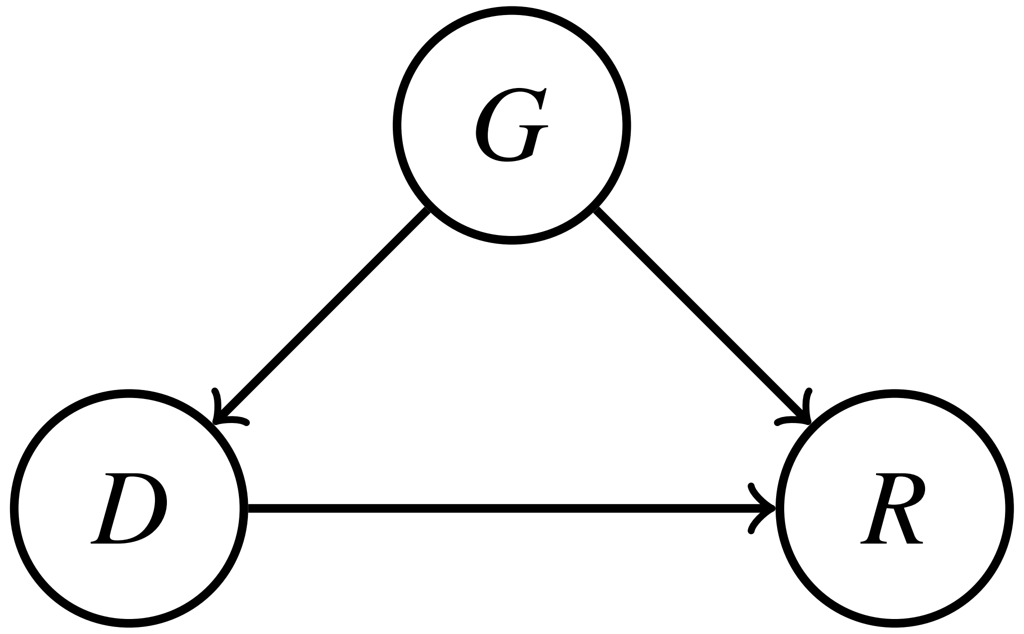
Simpson's paradox: Case 1

Gender as a confounder

- 700 patients w/ choice: take drug or not
- variable of interest: recovery rate
- also observed: **gender**

	Drug	No drug
Men	81 / 87 (93%)	234 / 270 (87%)
Women	192 / 263 (73%)	55 / 80 (68%)
Σ	273 / 350 (78%)	289 / 350 (83%)

causal relation
gender is
a **confound**



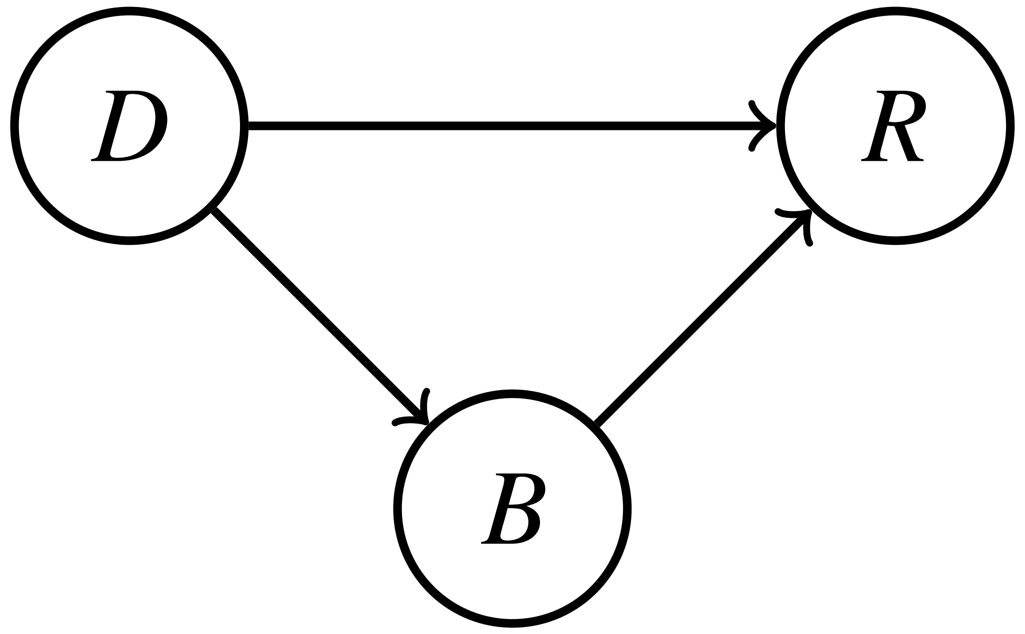
Case 2

Blood-pressure as a mediator

- same as case 1, but no gender info
- also observed:
  post-treatment blood pressure

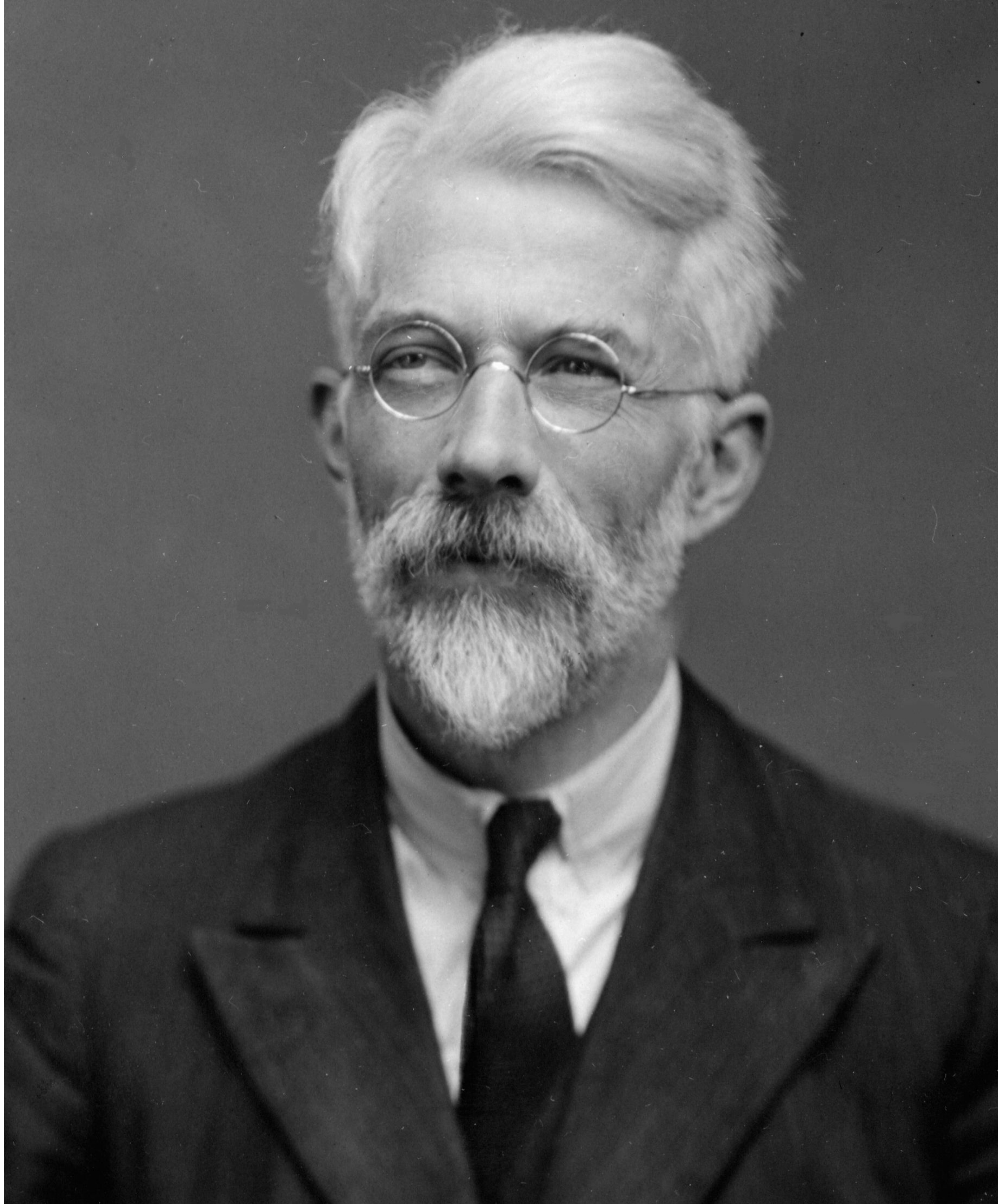
	Drug	No drug
Low BP	81 / 87 (93%)	234 / 270 (87%)
High BP	192 / 263 (73%)	55 / 80 (68%)
Σ	273 / 350 (78%)	289 / 350 (83%)

causal relation
blood pressure is
a **mediator**



“the object of statistical methods is the reduction of data”

Fisher (1922)



causal models

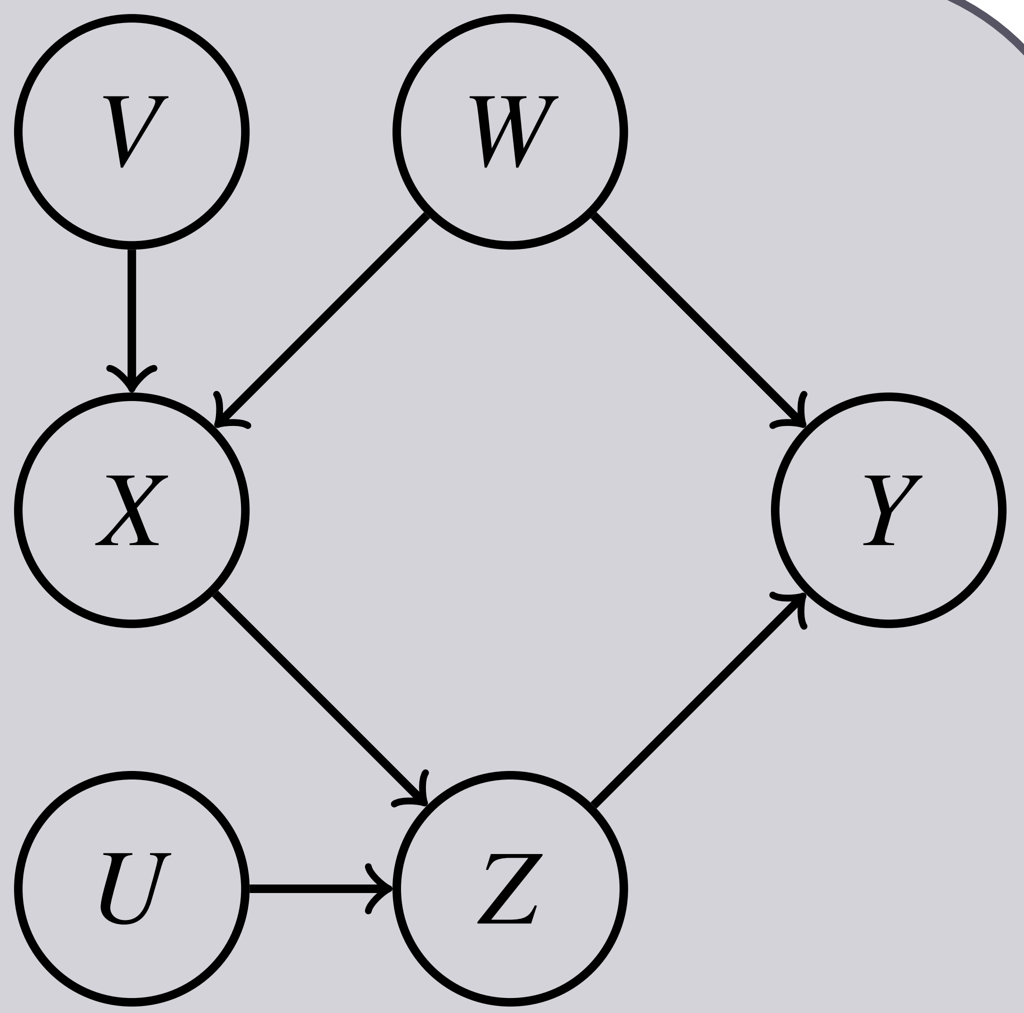
Causal models

intuitive, informal approach

"causal models represent the *mechanism* by which data were generated" (Pearl et. al 2016, p. 36)

- directed acyclic graph (DAG):
  - nodes are variables / events
  - edges indicate direct causal relationship
  - paths indicate (indirect) causal relationship
- beliefs in causal relationships constrain beliefs in stochastic dependency
  - no causal path => stochastic independence
  - single causal path from X to Y via Z => conditional stochastic dependence

example



- V and W are independent ($V \perp\!\!\!\perp W$)
- U and Y are independent conditional on Z ($U \perp\!\!\!\perp Y \mid Z$)
- …

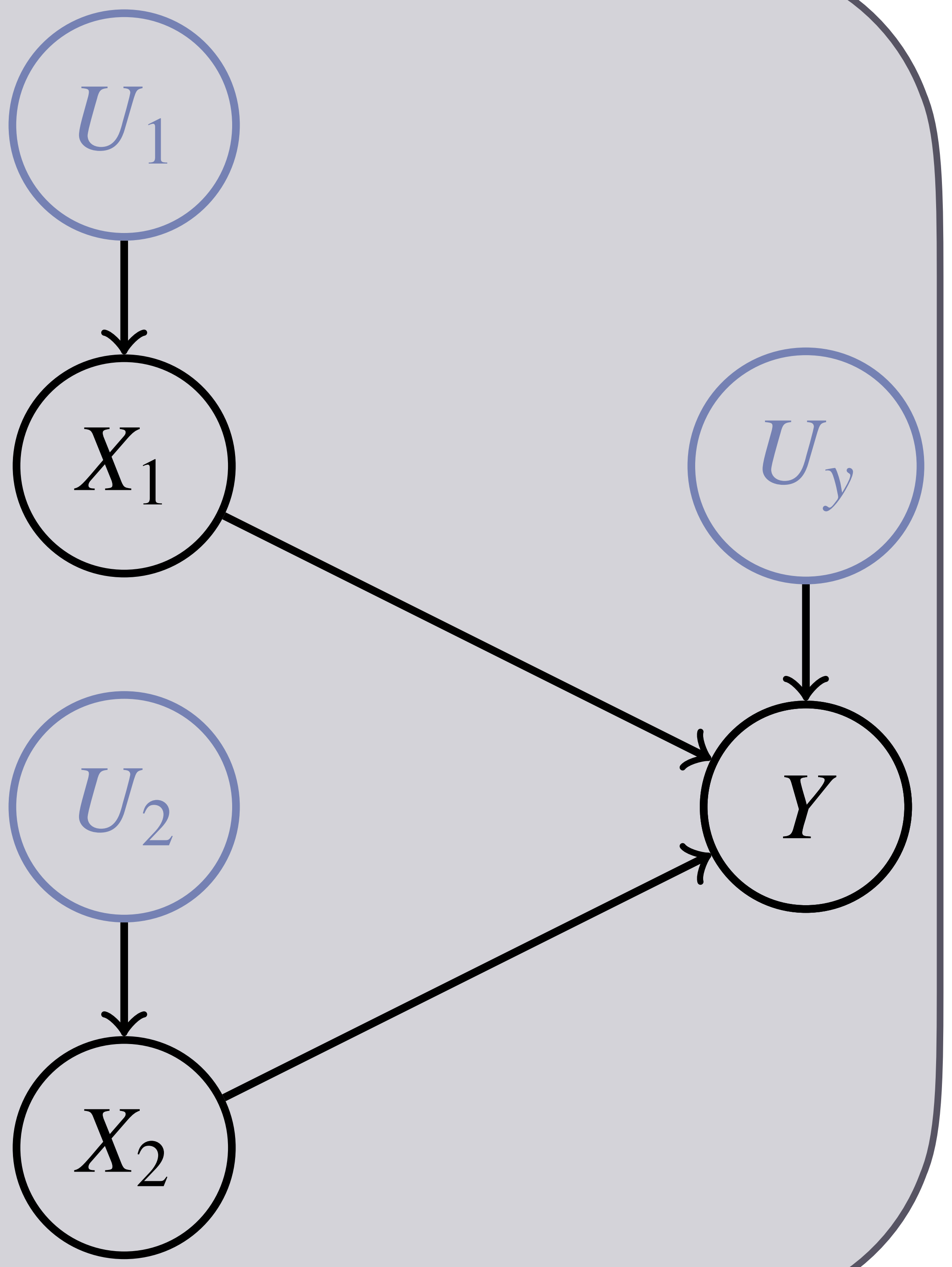
Excursion: Structural causal models

separating stochasticity and deterministic causal relations

- a structural causal model (SCM) consists of:
  - set U of exogenous variables
    - cause-less "first-movers" / root variables
    - introduce all randomness
  - set V of endogenous variables
  - set F of functions determining the value of some $v \in V$, based on the values of some subset of $U \times V$
    - all dependencies are deterministic
- advantage: can capture complex causal dependencies involving error terms

example

- exogenous variables are sampled from normal distributions
- $x_1 = u_1$
- $x_2 = u_2$
- $y = \beta_0 + \beta_1 x_1 + \beta + 2x_2 + u_y$

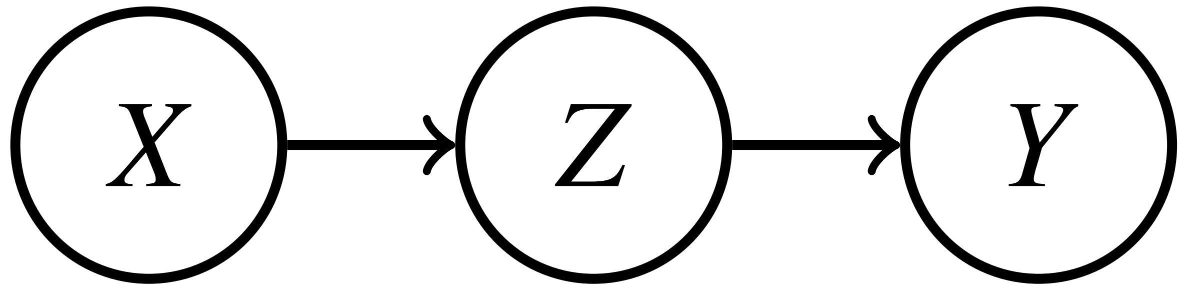


Pearl et al. (2016)

Elementary causal relationships

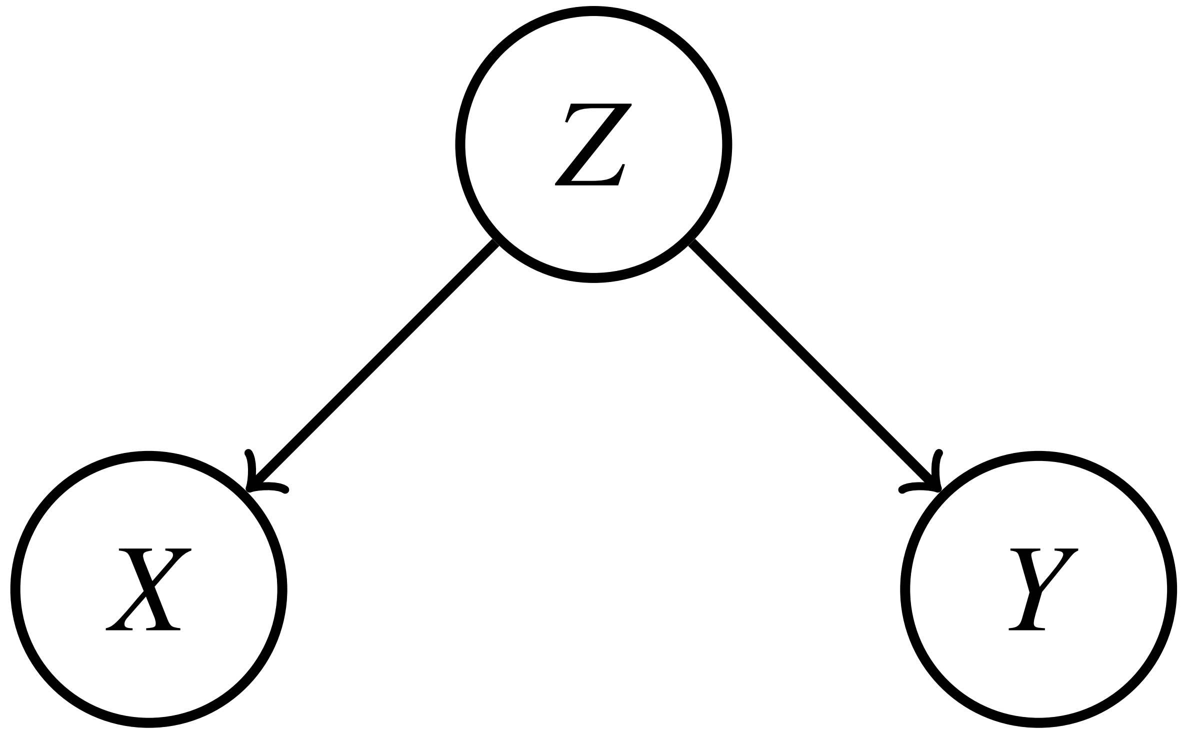
think: “conceptual atoms of complex causal graphs”

Chain



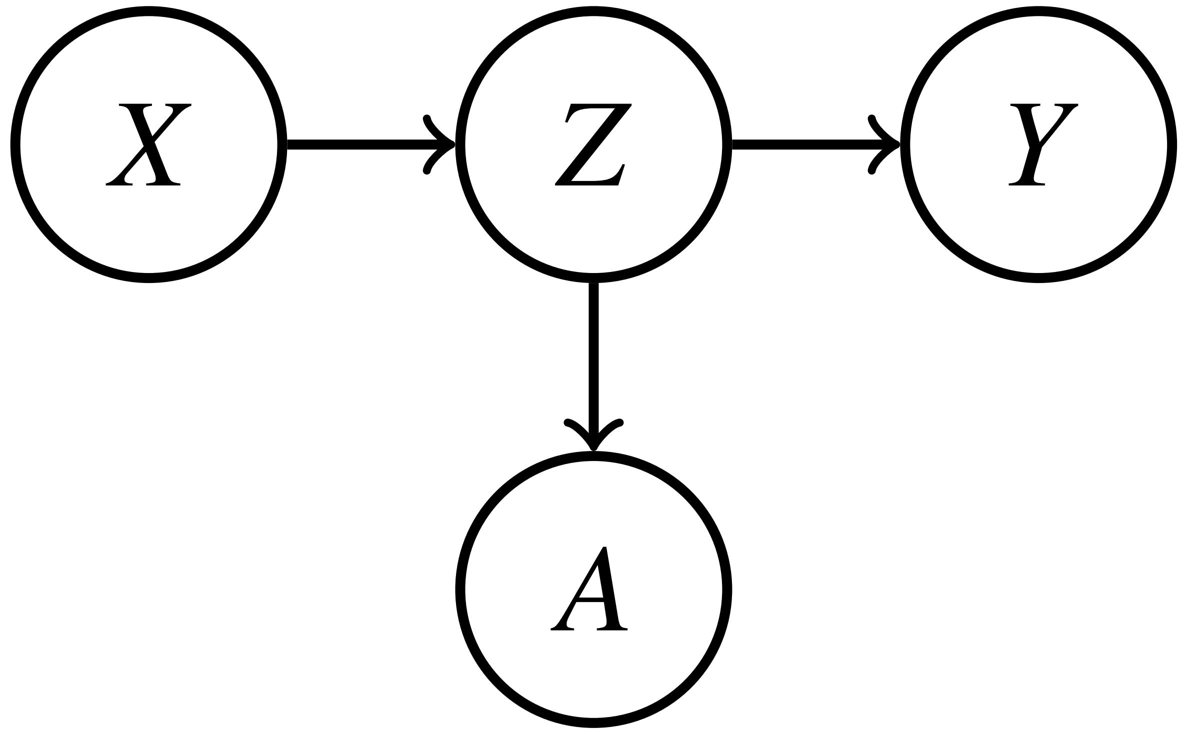
- X & Y independent conditional on Z

Fork



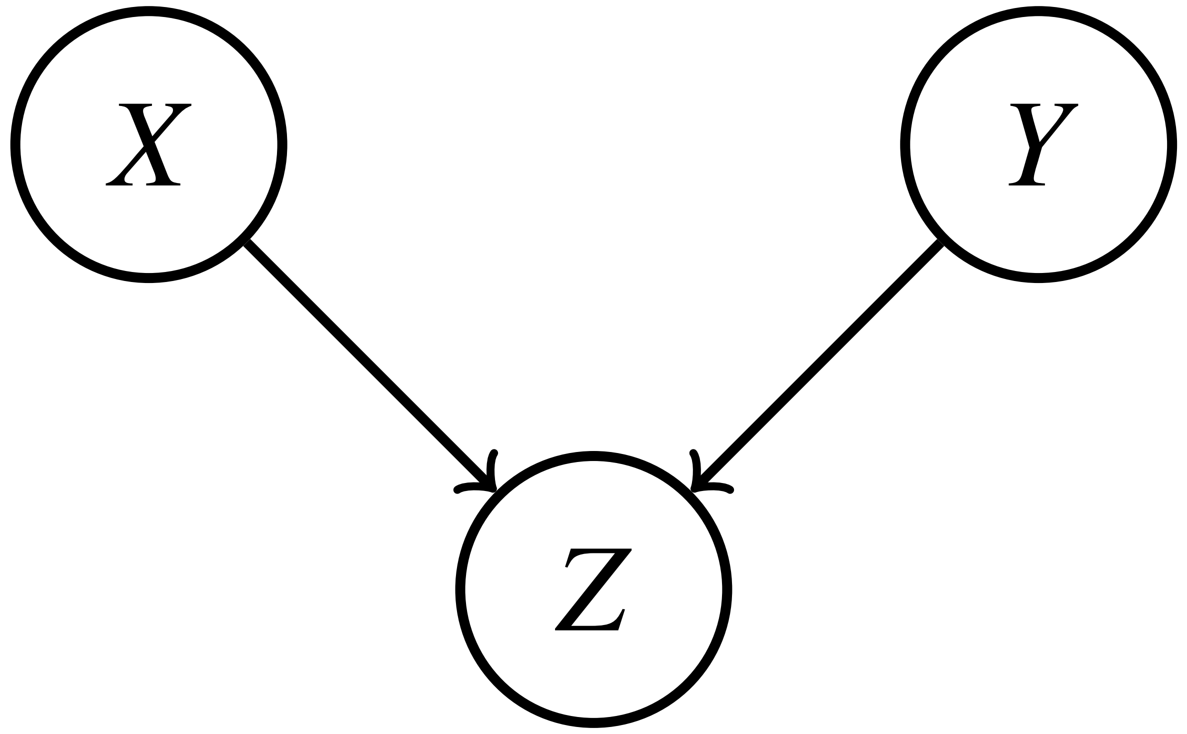
- X & Y stochastically dependent without direct causal relation

Descendant



- X & Y independent conditional on A, the more A provides information about Z

Collider



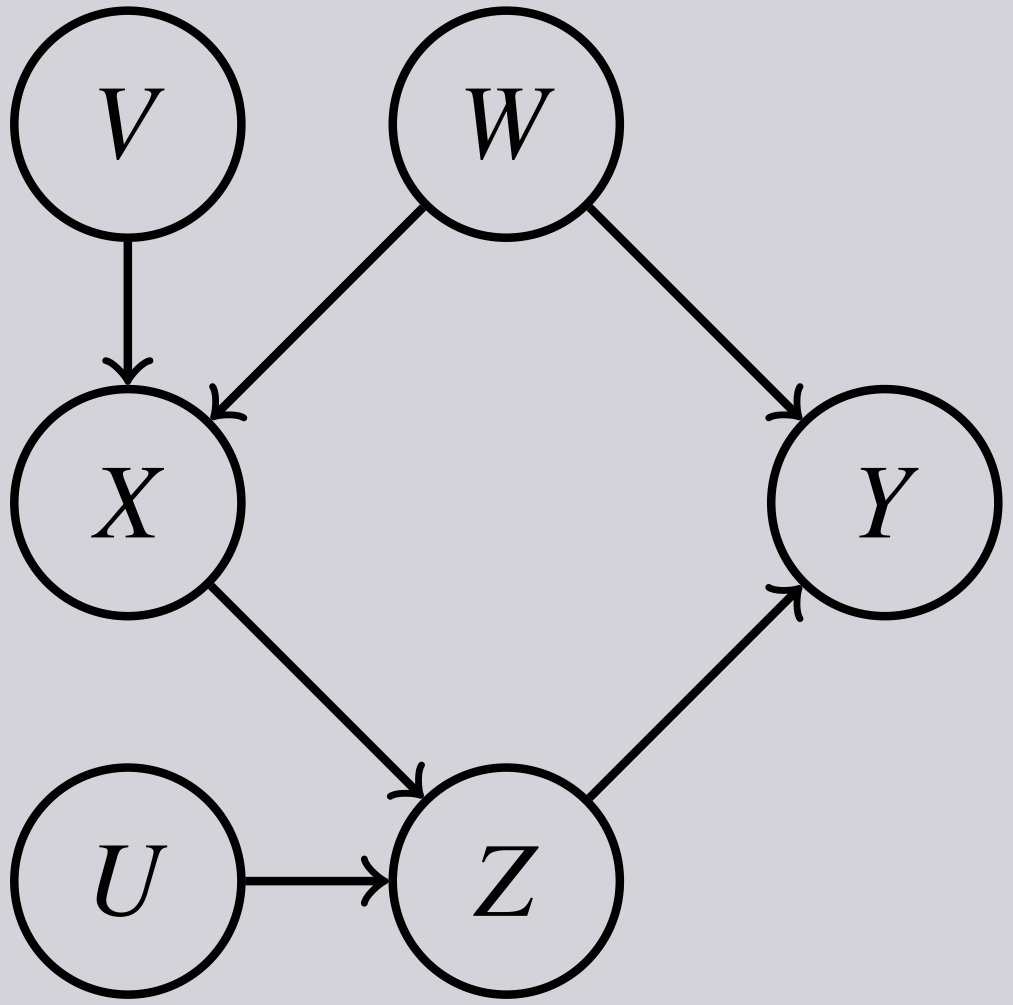
- X & Y independent, but dependent conditional on Z

d-separation

which nodes are (in-)dependent?

- fix a graph of causal relations
- each pair of nodes (X, Y) is **either:**
  - **d-separated**: X and Y are definitely independent
  - **d-connected**: X and Y are possibly dependent
    - they *are* dependent, except for fringe cases
- X and Y are **d-separated** if:
  - all paths between them are blocked
- a path is blocked if:
  - it passes through a collider

example



d-separated nodes

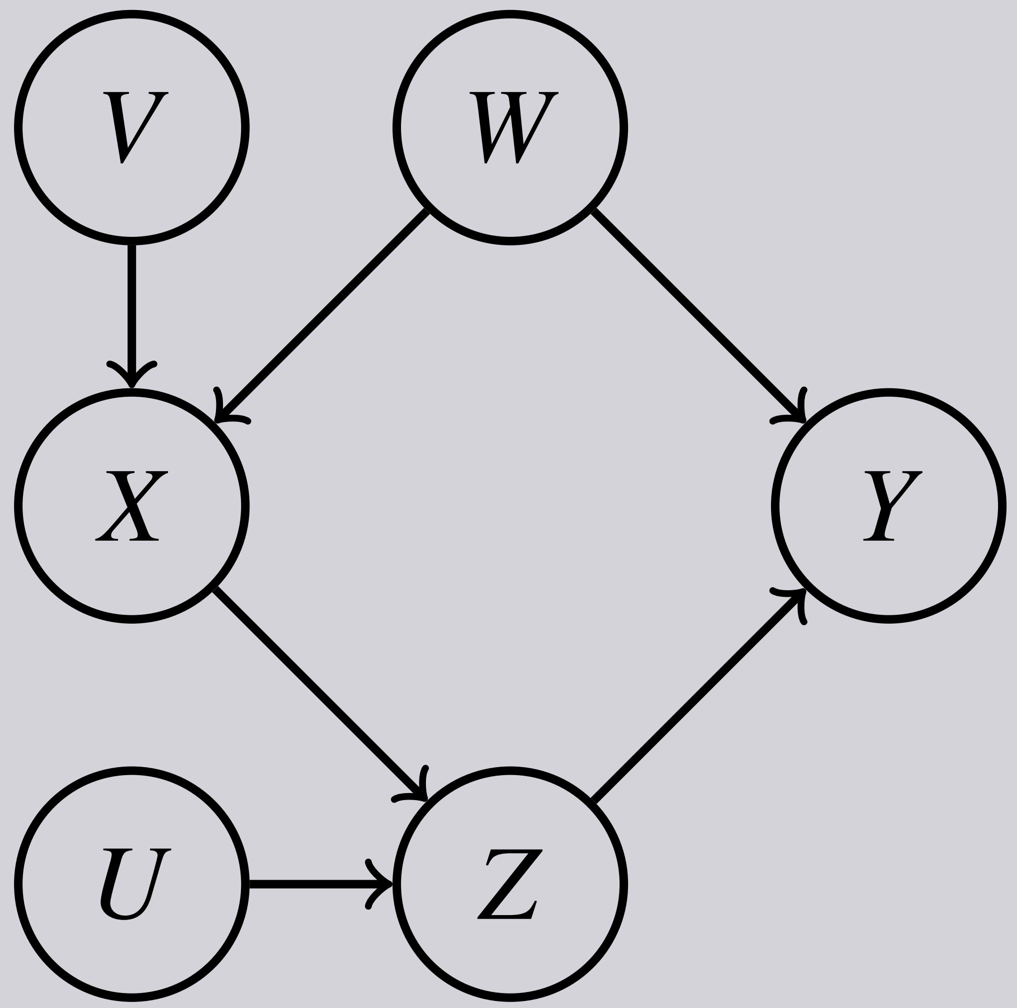
- $(U, V), (X, U), (V, W)$

conditional d-separation

which nodes are (in-)dependent given a set of observed variables?

- fix a set of nodes Z
- X and Y are **d-separated conditional on Z** if:
  - all paths between X and Y are blocked *simpliciter* or by some node in Z
- a path is **blocked (simpliciter)** if:
  - it passes through a collider
- a path is **blocked by a node Z** if:
  - Z is the middle node of a chain or fork
- a path is **unblocked by a node Z** if:
  - it is blocked w/o Z, and Z is a collider or a descendent of one

example



d-separated nodes

- $(U, V), (X, U), (V, W)$

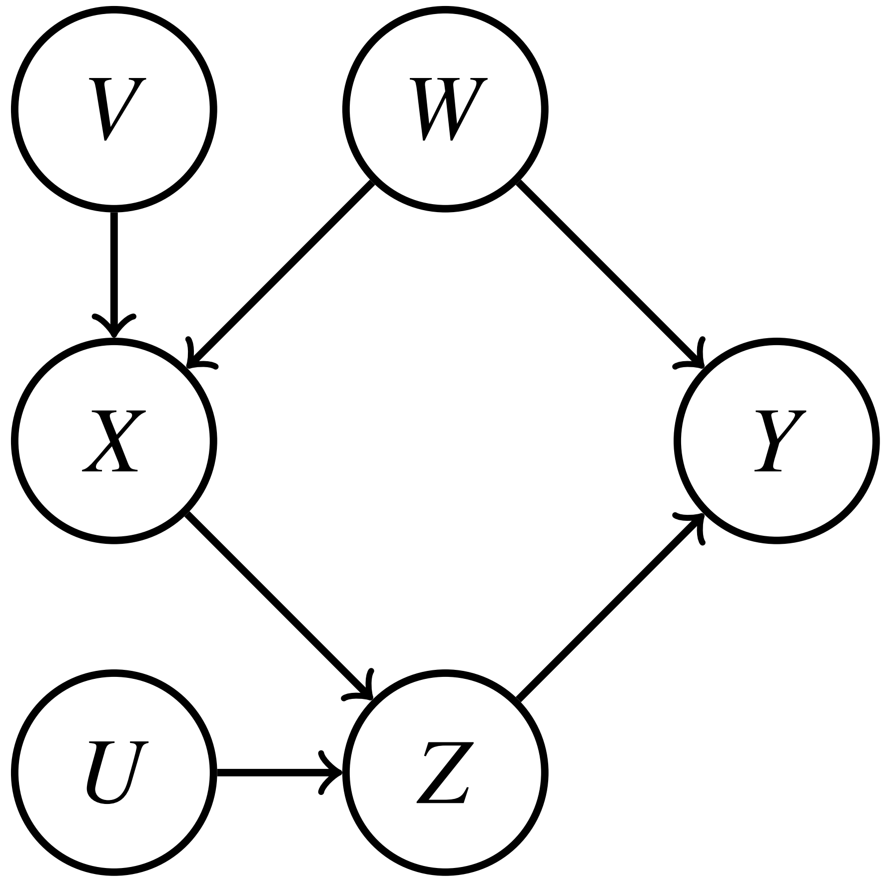
X and Y d-separated conditional on:

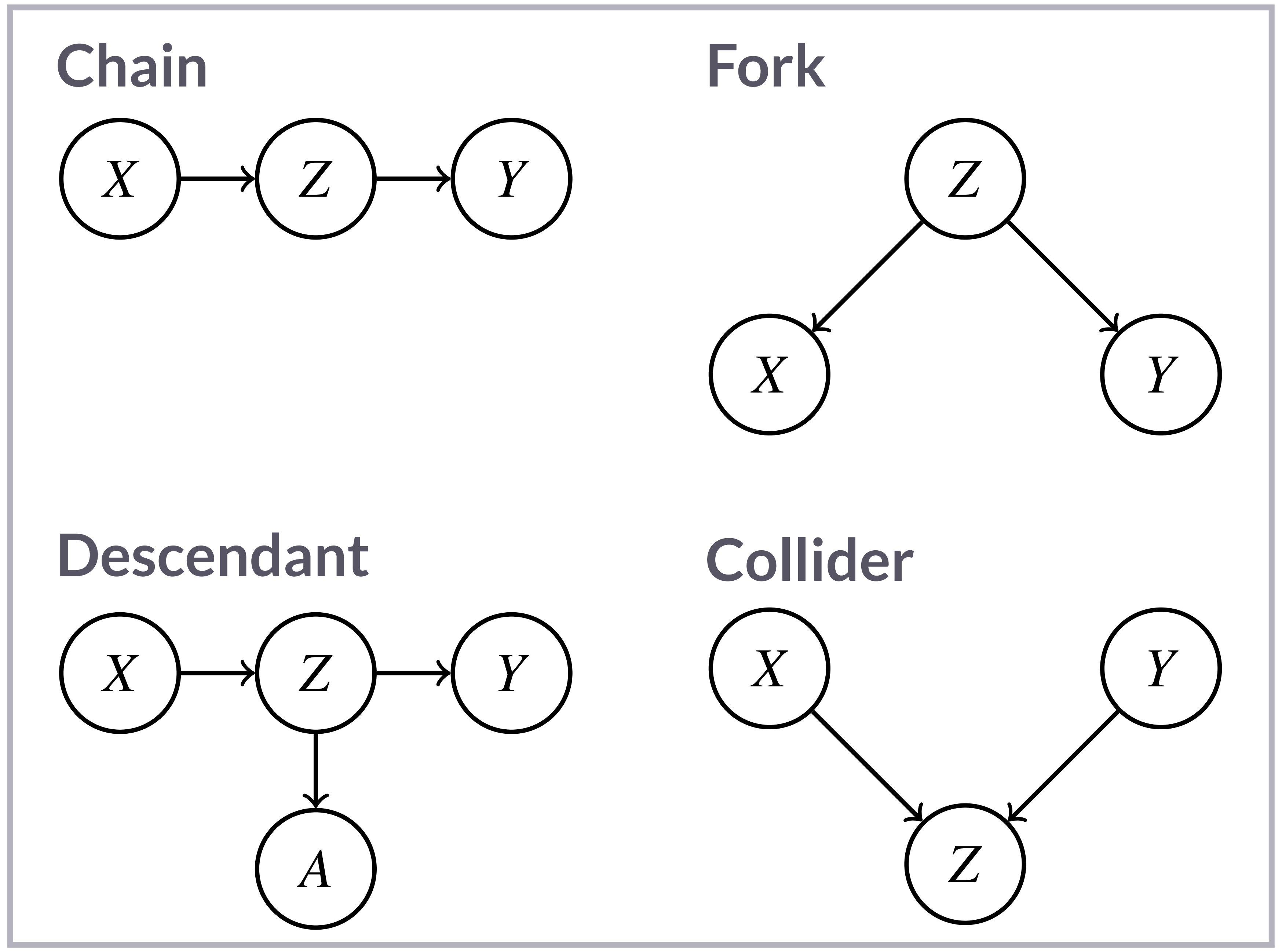
- $\{W, Z\}, \{U, W, Z\}, \{V, W, Z\}, \{U, V, W, Z\}$

causal models

summary

- capture (modellers') assumptions about causal relations
- entail assumptions about stochastic (in-)dependencies
- four elementary building blocks
  - chains
  - forks
  - descendants
  - colliders
- **d-separation:** method of determining (conditional) stochastic dependencies in any given graph





interventions

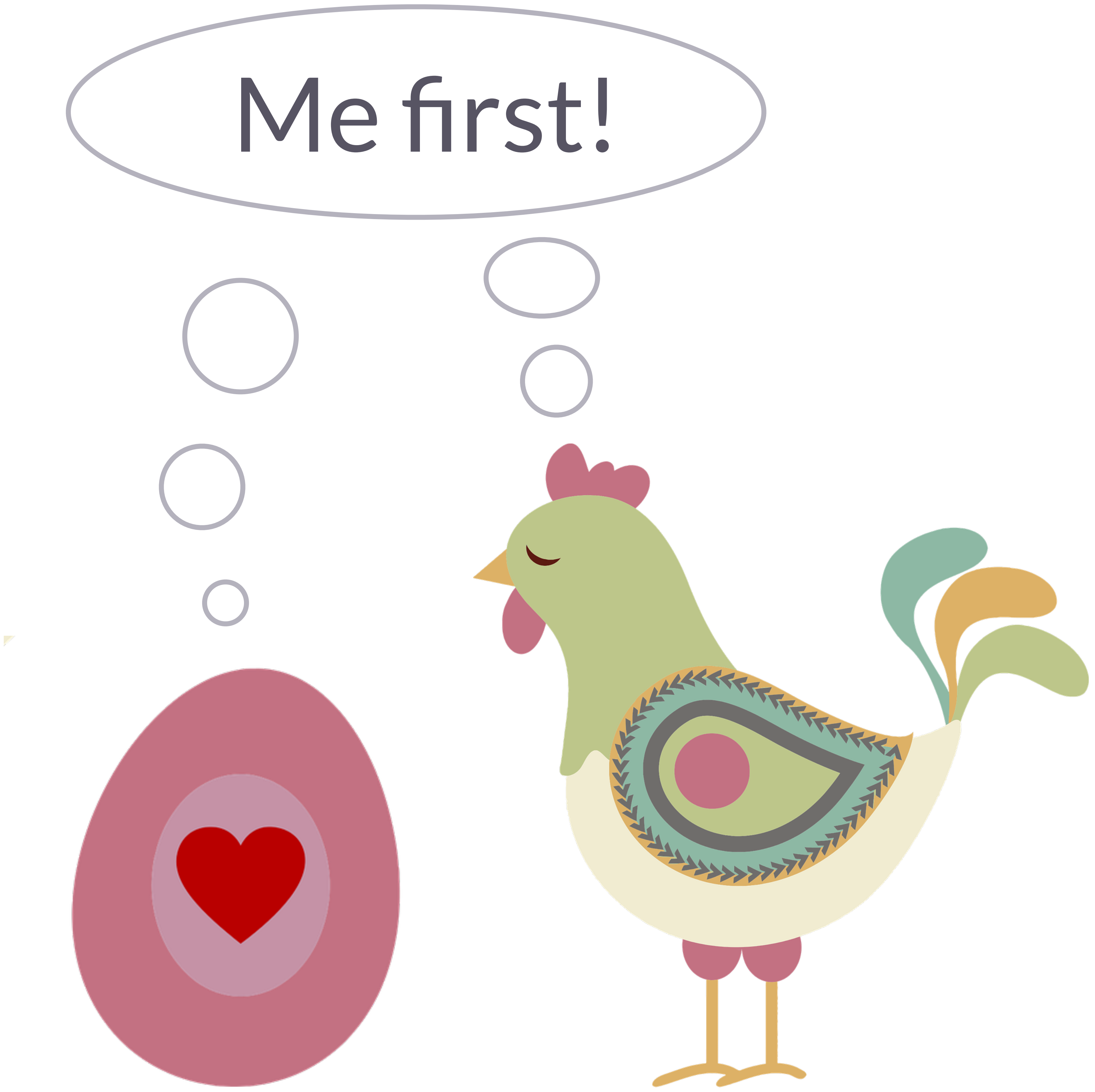
Interventions

Causal intuitions beyond stochastic dependencies

- intuitions about causal dependencies go beyond (intuitions about) stochastic dependencies
- example: different causal DAGs can have the same stochastic dependencies

$$X \rightarrow Y \qquad\qquad X \leftarrow Y$$

- differentiated by intuitions about **interventions**
  - changing X (all else equal) causes changes in Y only if Y is causally downstream of X
  - standard approach to intervention (not always possible):
    - randomized controlled trails (RCTs)
- goal (sometimes achievable, but not always):
  - capture effects of intervention without intervening

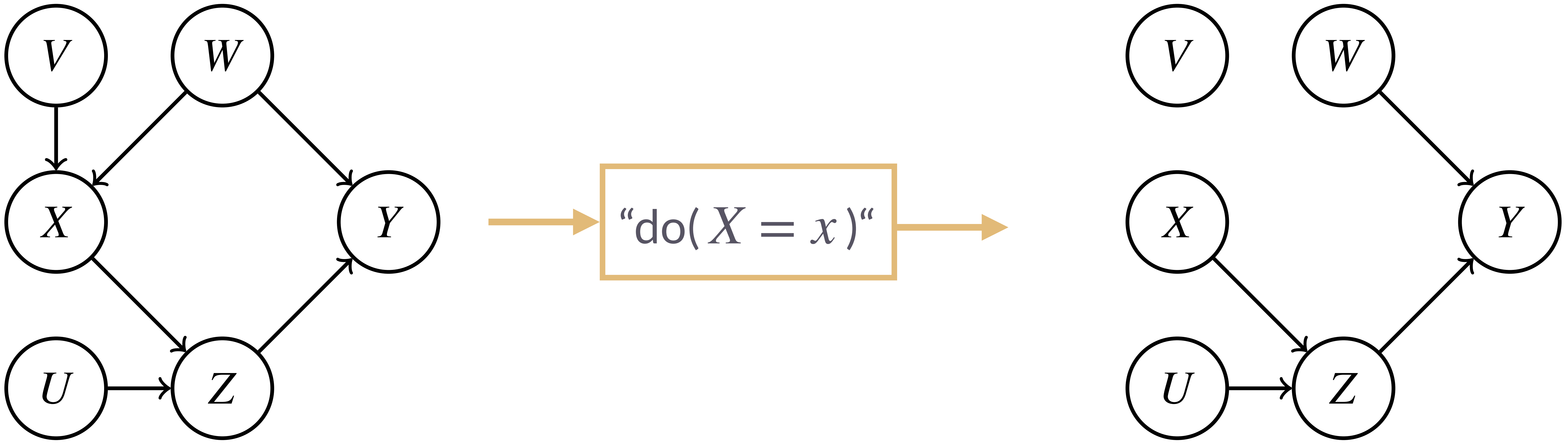


do-calculus

Formalizing effects of interventions

- new notation $P(Y = y \mid do(X = x))$:
  - probability of $Y = y$ after intervening the causal flow by setting $X = x$
- intervening = pruning:
  - **"doing $X = x$" => remove all arrows pointing towards X**

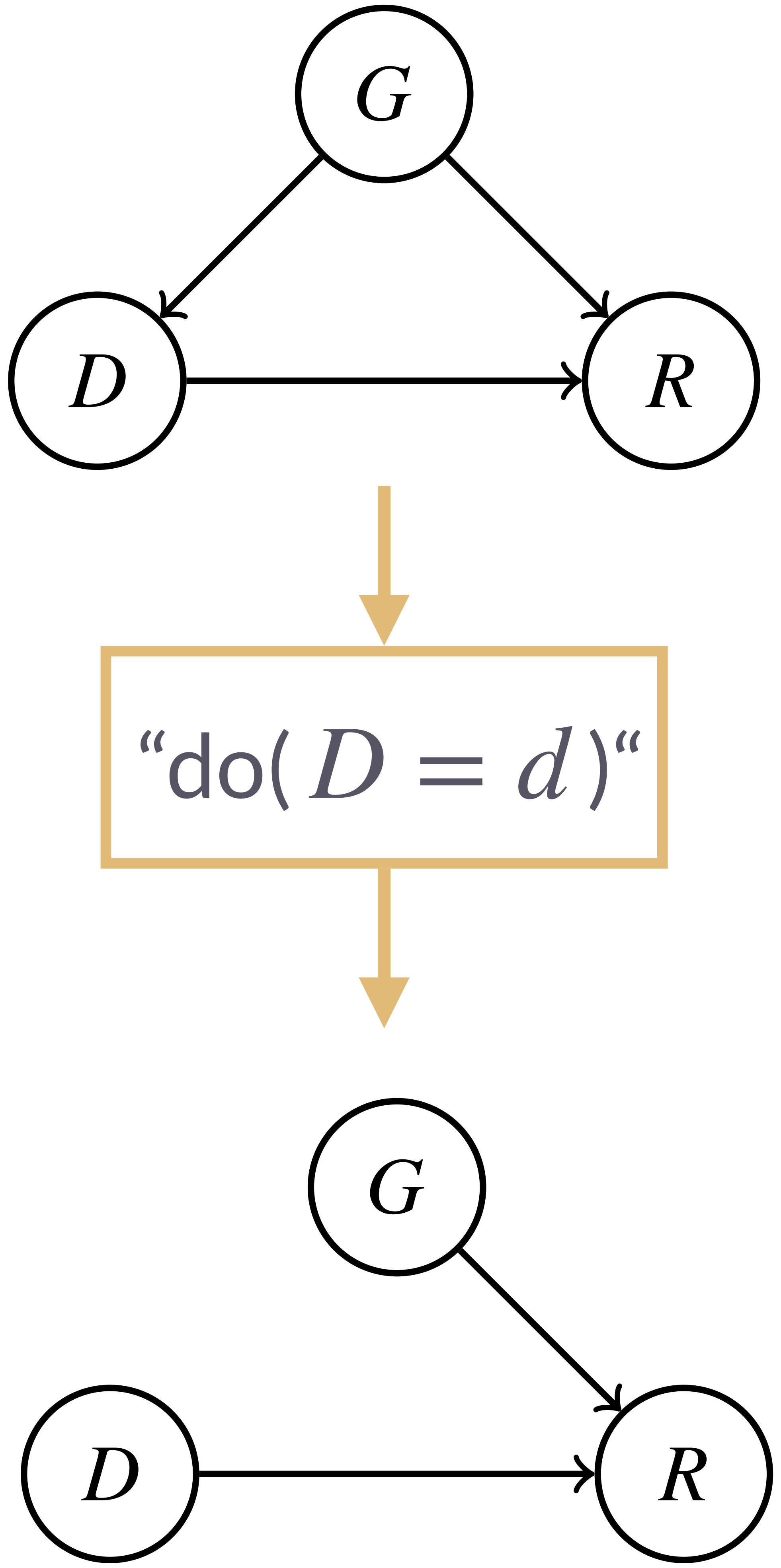
Causal dependencies **Updated graph**



NB: influence of X on Y now passes only via Z, not the confounder W

Example: Simpson’s paradox

Case 1: gender as a confounder



$$P\left(R = r \mid do(D = d)\right)$$

$$= P^*\left(R = r \mid D = d\right) \quad \text{[by definition]}$$

$$= \sum_g P^*\left(R = r \mid D = d, G = g\right)\; P^*\left(G = g \mid D = d\right) \quad \text{[rules of prob.]}$$

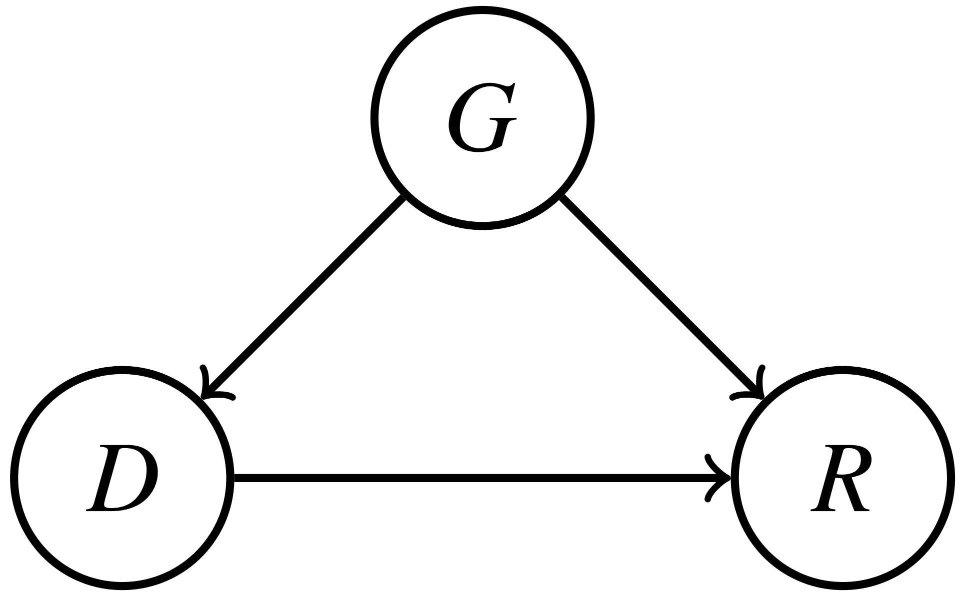
$$= \sum_g P^*\left(R = r \mid D = d, G = g\right)\; P^*\left(G = g\right) \quad \text{[independence]}$$

$$= \sum_g P\left(R = r \mid D = d, G = g\right)\; P\left(G = g\right) \quad \text{[not affected by “do”]}$$

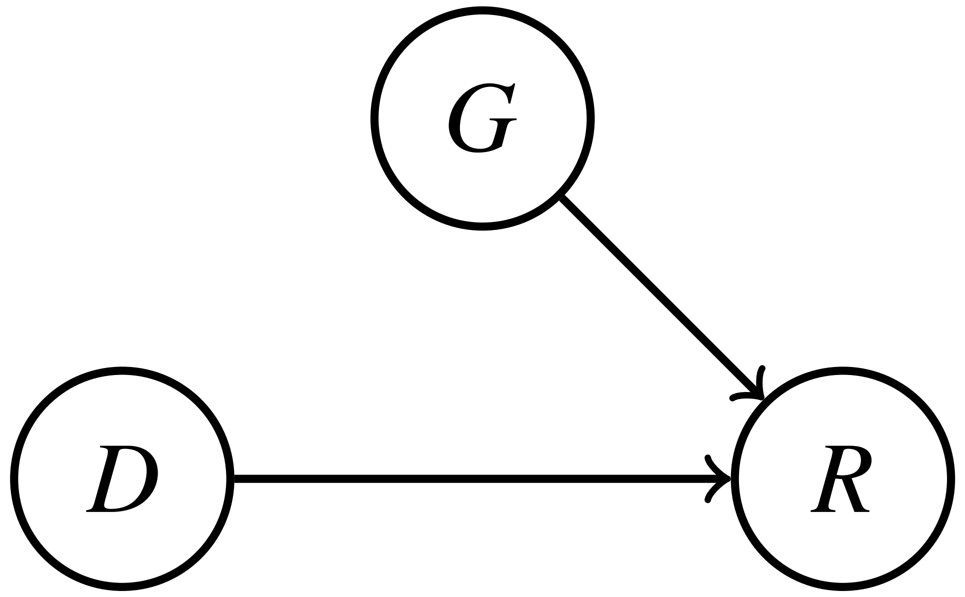
!!! It is possible to express effects of “do”-intervention in terms of observational probabilities alone!

Example: Simpson's paradox

Case 1: gender as a confounder



"do($D = d$)"



$$P\left(R = r \mid do(D = d)\right)$$

$$= \sum_{g} P\left(R = r \mid D = d, G = g\right) \; P\left(G = g\right)$$

	Drug	No drug
Men	81 / 87 (93%)	234 / 270 (87%)
Women	192 / 263 (73%)	55 / 80 (68%)
Σ	273 / 350 (78%)	289 / 350 (83%)

$$P\left(R = 1 \mid do(D = 1)\right) \approx 0.83$$

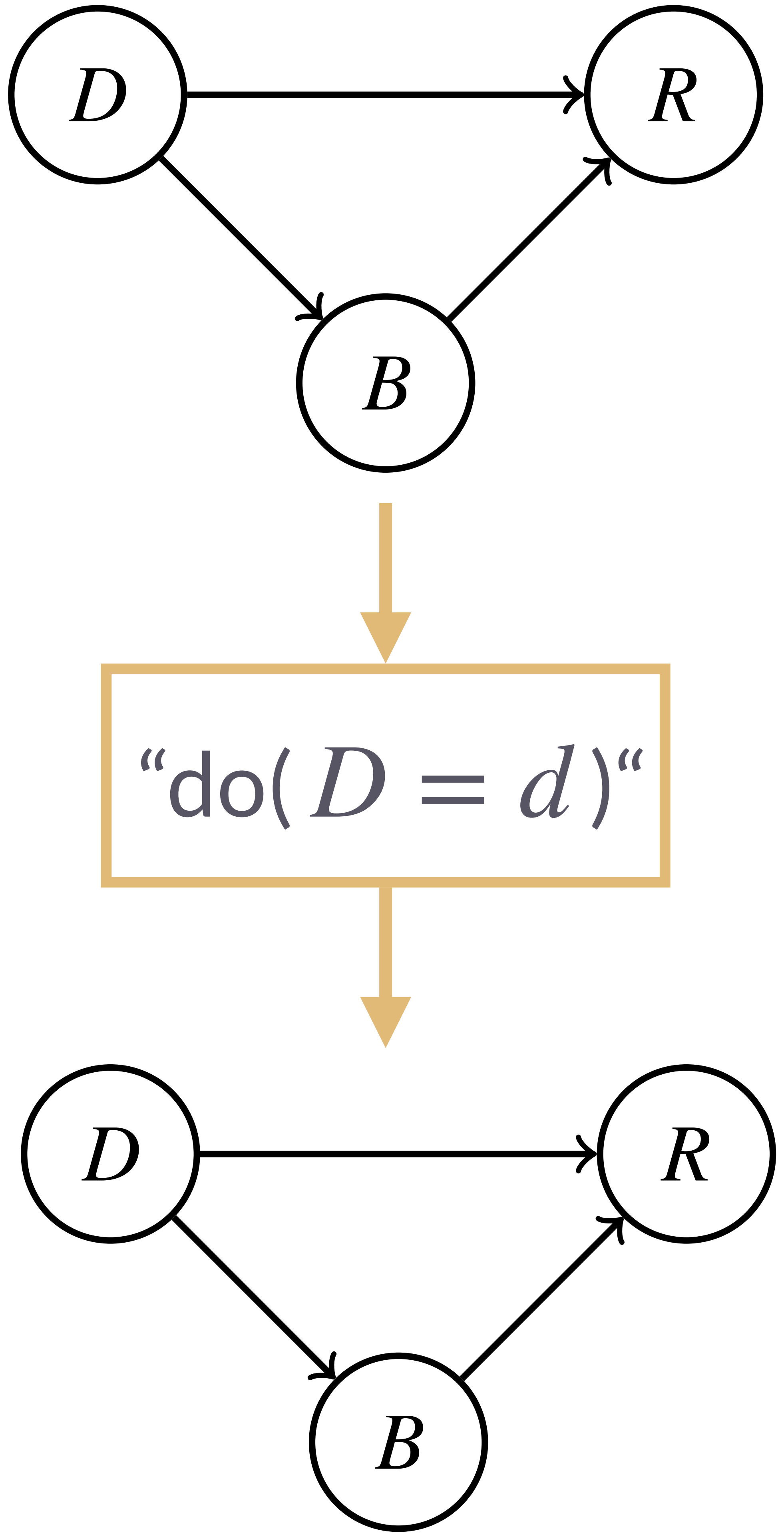
$$P\left(R = 1 \mid do(D = 0)\right) \approx 0.78$$

ML-estimate of **causal effect**

$$0.83 - 0.78 = 0.05$$

Example: Simpson's paradox

Case 2: blood pressure as a mediator



$$P\left(R = r \mid do(D = d)\right)$$

$$= P^*\left(R = r \mid D = d\right)$$ [by definition]

$$= \sum_b P^*\left(R = r \mid D = d, B = b\right)\; P^*\left(B = b \mid D = d\right)$$ [rules of prob.]

$$\cancel{= \sum_b P^*\left(R = r \mid D = d, B = b\right)\; P^*\left(B = b\right)}$$ ~~[independence]~~

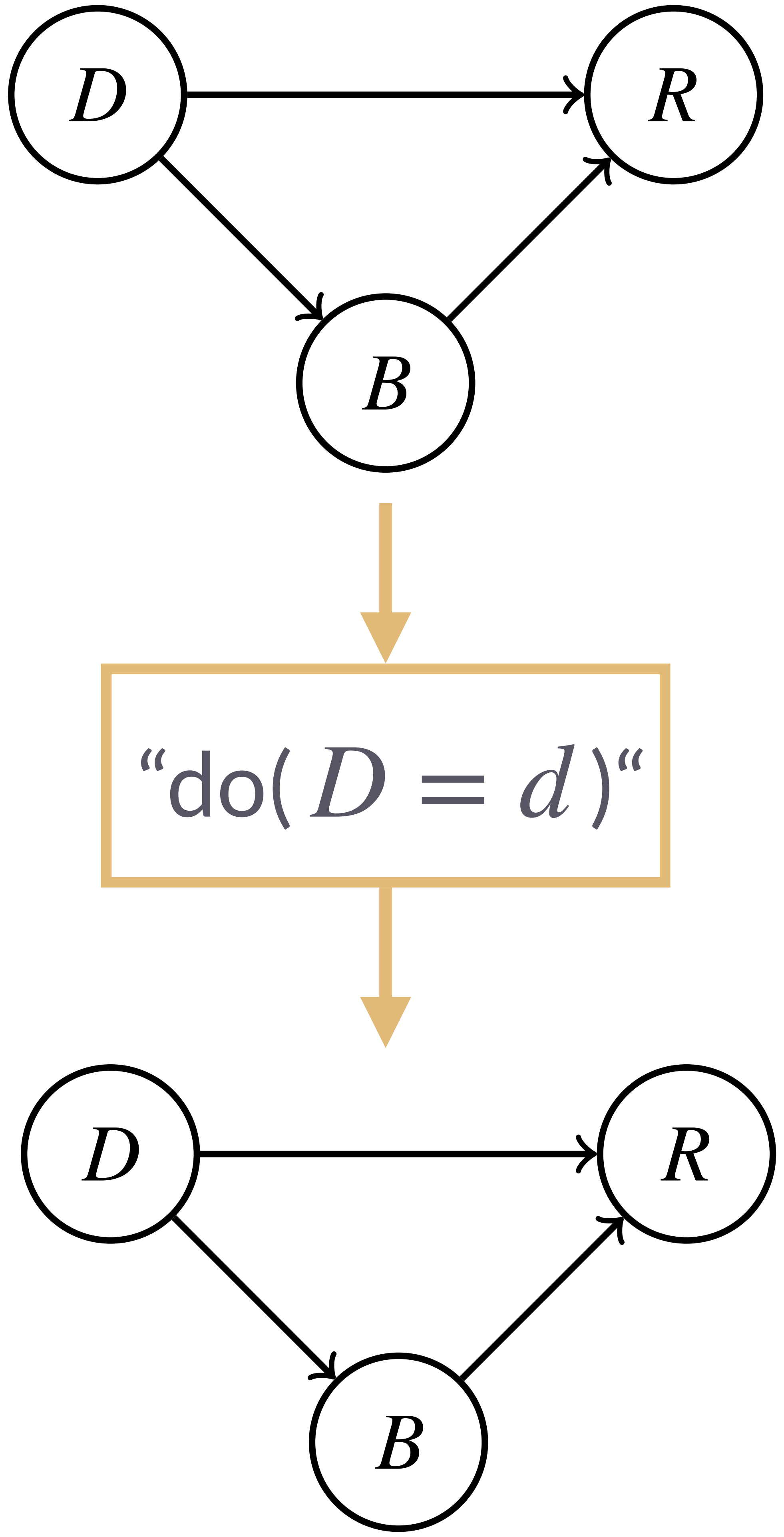
$$= \sum_b P\left(R = r \mid D = d, B = b\right)\; P\left(B = b \mid D = d\right)$$

$$= P\left(R = r \mid D = d\right)$$

Here "do"-intervention does not "create independency".
And it is not necessary at all!

Example: Simpson's paradox

Case 2: blood pressures as a mediator



$$P\left(R = r \mid do(D = d)\right)$$
$$= \sum_{b} P\left(R = r \mid D = d, G = g\right)\; P\left(G = g \mid B = b\right)$$

	Drug	No drug
Low BP	81 / 87 (93%)	234 / 270 (87%)
High BP	192 / 263 (73%)	55 / 80 (68%)
Σ	273 / 350 (78%)	289 / 350 (83%)

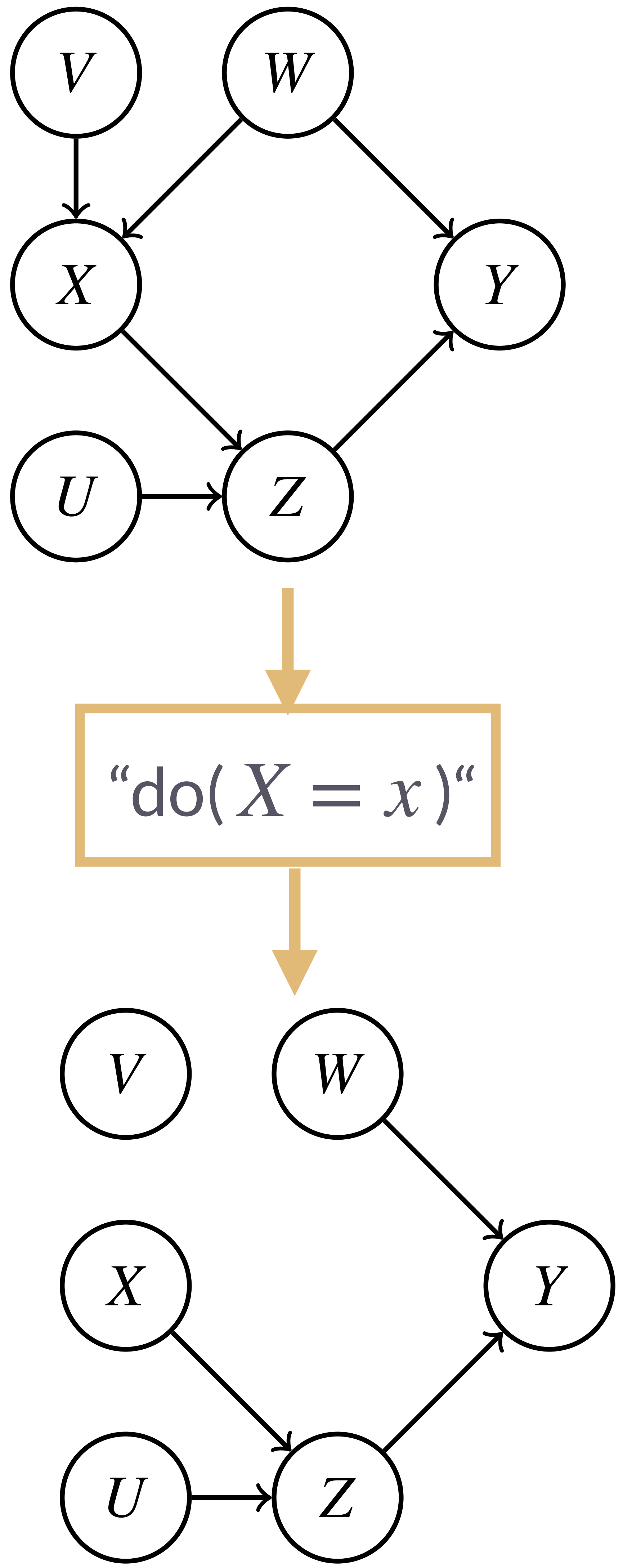
$$P\left(R = 1 \mid do(D = 1)\right) \approx 0.78$$
$$P\left(R = 1 \mid do(D = 0)\right) \approx 0.83$$

ML-estimate of
causal effect
$0.78 - 0.83 = -0.05$

Intervention

summary

- causal intuitions go beyond stochastic dependence, and imply intuitions about **interventions**
- intervening = pruning:
  - “doing $X = x$“ entails removing all arrows pointing towards X
- new formal notion: $P(Y = y \mid do(X = x))$
- sometimes we can express $P(Y = y \mid do(X = x))$ in terms of “normal” conditional probabilities; sometimes we cannot
- follow-up question: when and how can we eliminate “*do(X)*”?



eliminating “do X”

or: blocking backdoors & taking front-doors

Backdoor criterion

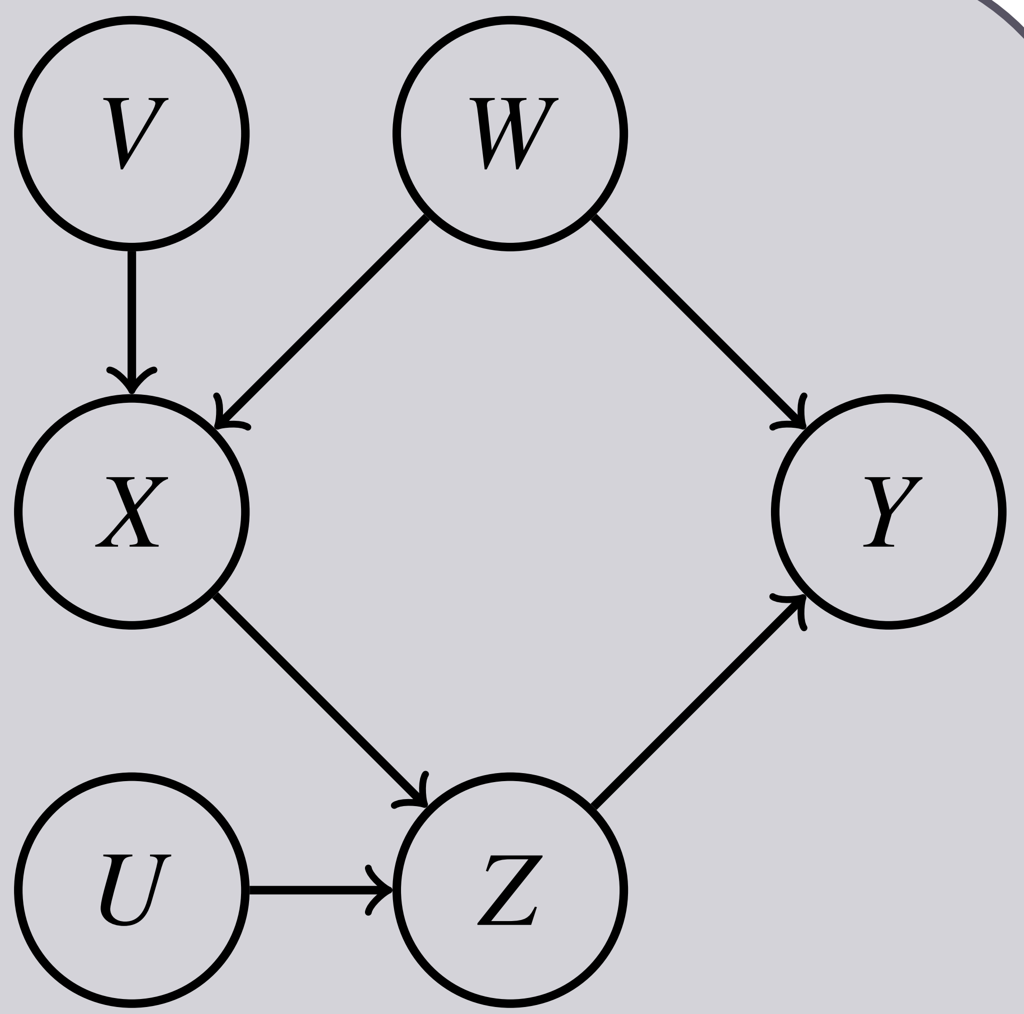
Definition

A set of nodes Z satisfies the BDC relative to a pair of nodes (X, Y) if

(1) Z contains no (direct or indirect) descendants of X

(2) for every path from X to Y that is unblocked (*simpliciter*) and contains an arrow into X, there is a node in Z which blocks this path

example



sets satisfying BDC for (X, Y)

- $\{W\}, \{U, W\}, \{V, W\}, \{U, V, W\}$

BDC theorem

If set Z satisfies the BDC for (X, Y), we can use the **adjustment formula**:

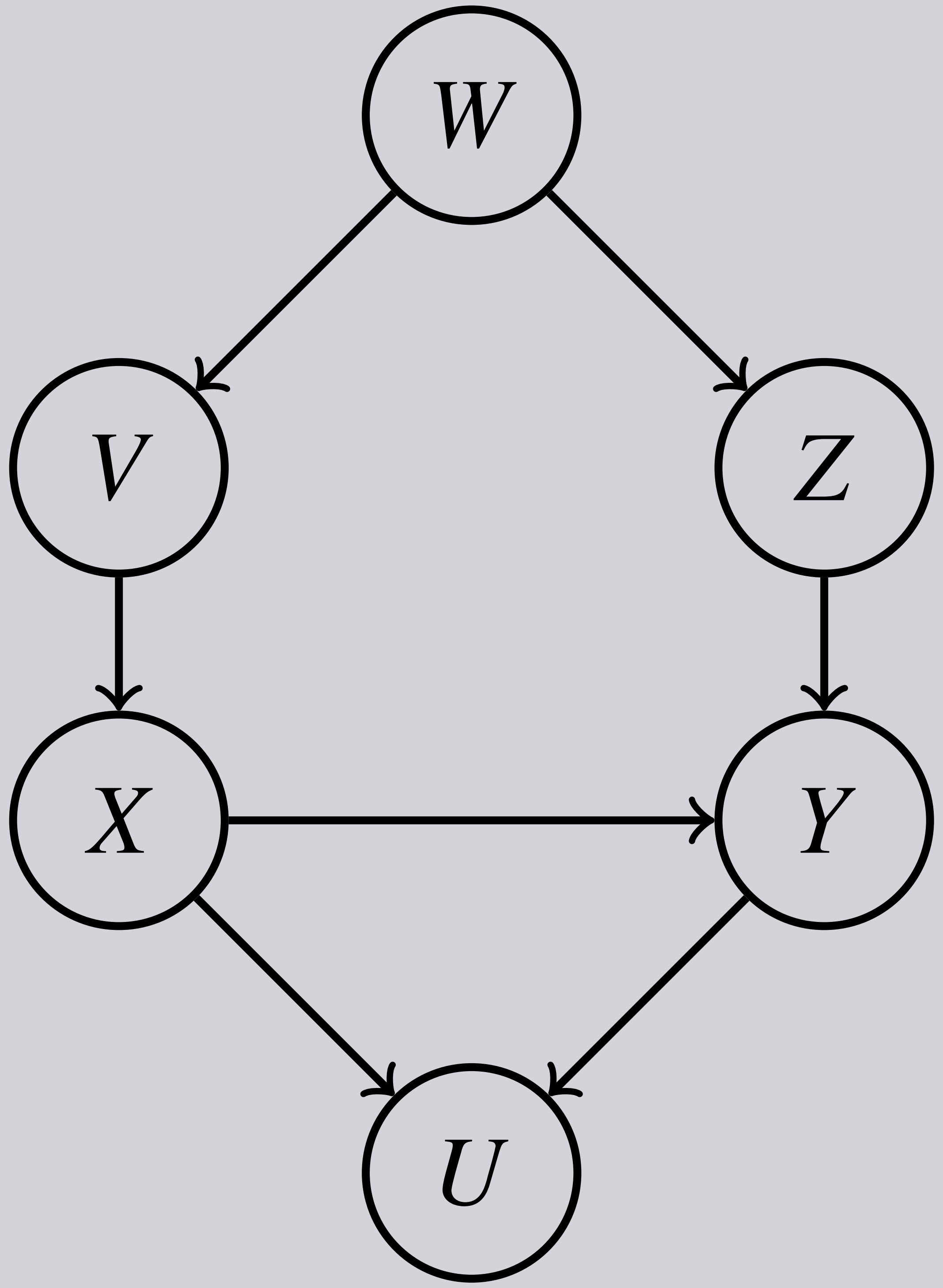
$$P\left(Y = y \mid do(X = x)\right) = \sum_z P\left(Y = y \mid X = x, Z = z\right) \; P(Z = z)$$

Usefulness of BDC

- trivial set satisfying BDC:
  - all parents of X
- minimal trivial set satisfying BDC:
  - all parents of X on a path to Y
- BDC is most useful if parents are unobserved or unobservable
- moreover: BDC equation can be used to test causal assumptions

  [causal model selection]

example



if V is unavailable we can instead condition on W, Z, or both

Front-door criterion

motivating example

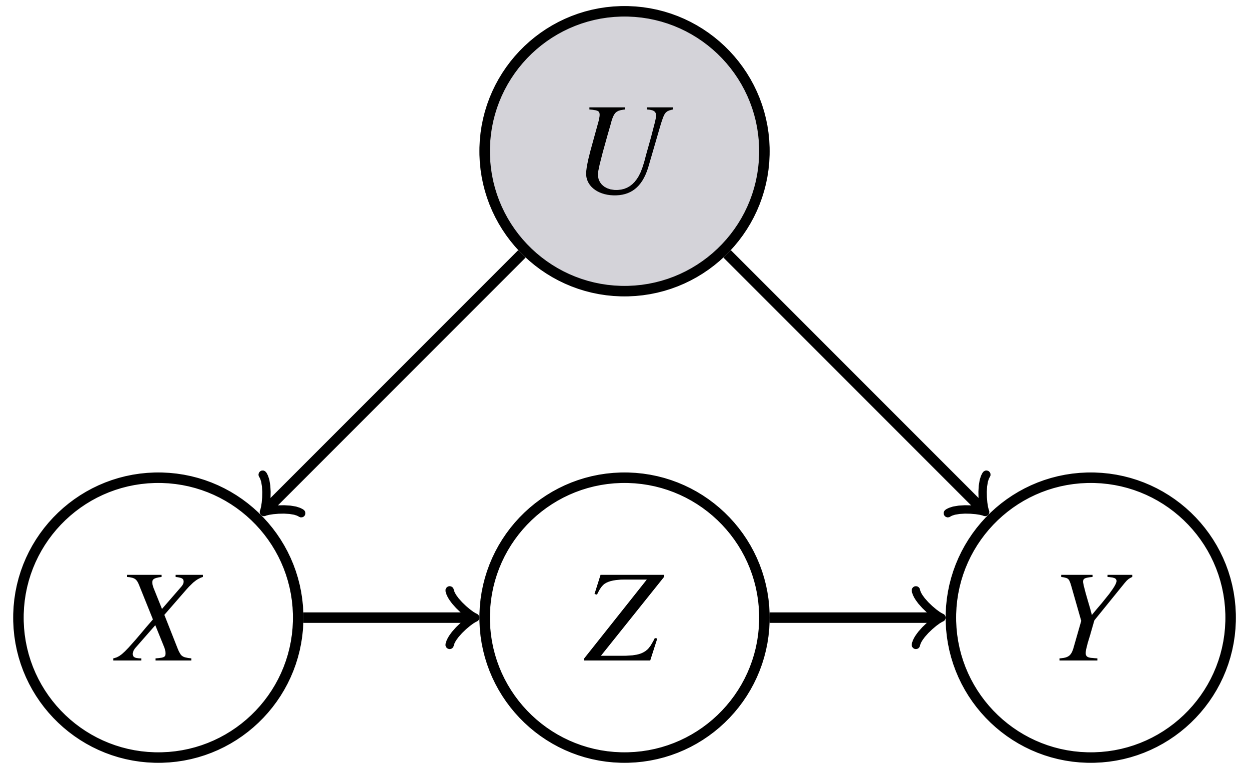
- causal DAG as shown on left
  - X, Y and Z are observed
  - U is not observed or unobservable
- we want to know $P\left(Y = y \mid do(X = x)\right)$
- BDC is not applicable directly (b/c U is not observed)
- but we can use BDC adjustment to compute:

(1) $P\left(Z = z \mid do(X = x)\right) = P\left(Z = z \mid X = x\right)$

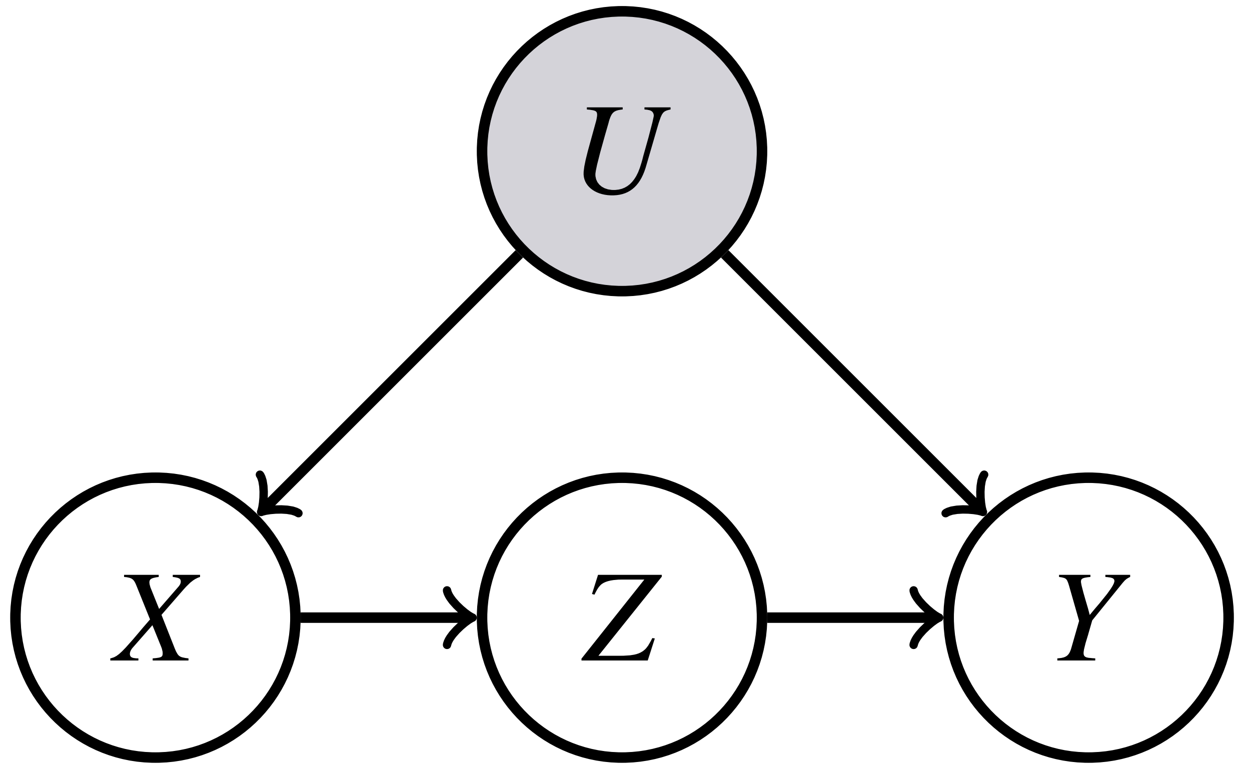
(2) $P\left(Y = y \mid do(Z = z)\right) = \sum_{x'} P\left(Y = y \mid Z = z, X = x'\right) \; P(X = x')$

- so, we combine this to **bang down the front-door**:

(3) $P\left(Y = y \mid do(X = x)\right) = \sum_{z} P(Z = z \mid X = x) \sum_{x'} P\left(Y = y \mid Z = z, X = x'\right) \; P(X = x')$



Front-door criterion



Definition

A set of nodes Z satisfies the FDC relative to a pair of nodes (X, Y) if

(1) Z intercepts all directed paths from X to Y

(2) there is no unblocked path from Z to X

(3) all backdoor paths from Z to Y are blocked by X

FDC theorem

If set Z satisfies the FDC for (X, Y) and $P(x, z) > 0$, we can use the **FDC-adjustment formula**:

$$P\left(Y = y \mid do(X = x)\right) = \sum_{z} P(Z = z \mid X = x) \sum_{x'} P\left(Y = y \mid Z = z, X = x'\right)\ P(X = x')$$

“do”-elimination
summary

- there are two standard criteria that identify variables so that we can “marginalize out the ‘do’”-operators
  - **backdoor criterion**: block all backdoor paths
  - **front-door criterion**: bang down the front door (and leave no traces)
- BDC and FDC cover many cases
- there are cases where “do”-elimination is possible which are *not* covered by either BDC or FDC
  - for these we need to apply the “do”-calculus and think-in-context 😰

causal effects

Causal effects

what happens when we "$do(X = x)$"?

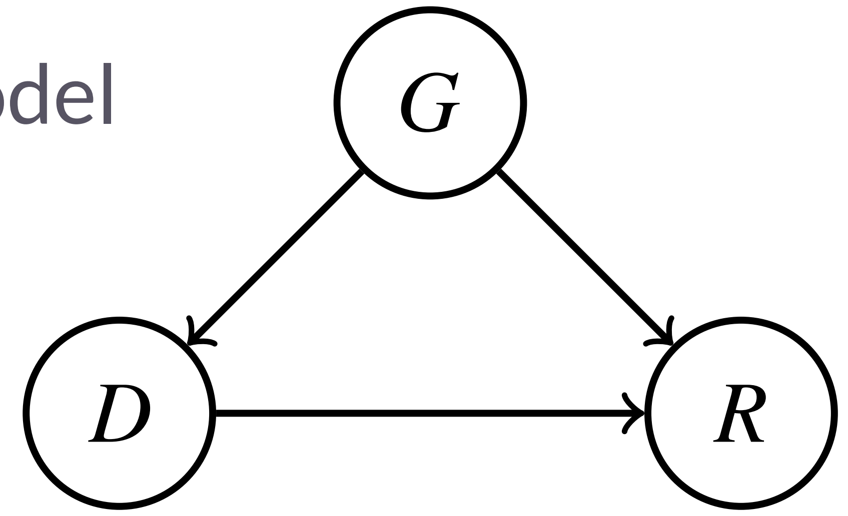
- "causal effect" is an elusive cluster concept!
  - depends on type of relevant variables
  - depends on what we care about
- general idea:
  - how much influence does $do(X = x)$ have on the distribution over Y
- causal effect of binary X on binary Y:

$$P(Y = 1 \mid do(X = 1)) - P(Y = 1 \mid do(X = 0))$$

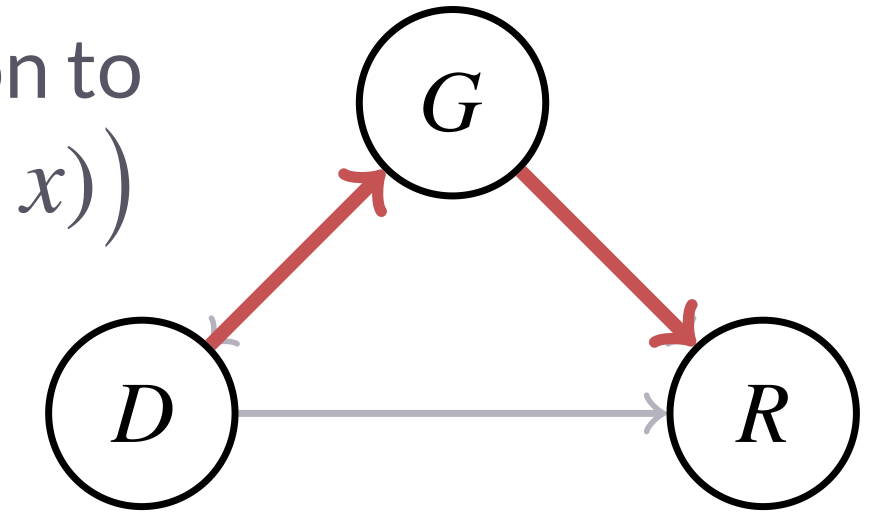
- causal effect of continuous X on continuous Y:

$$\mathbb{E}\left(Y^{|do(X=1)}\right) - \mathbb{E}\left(Y^{|do(X=0)}\right)$$

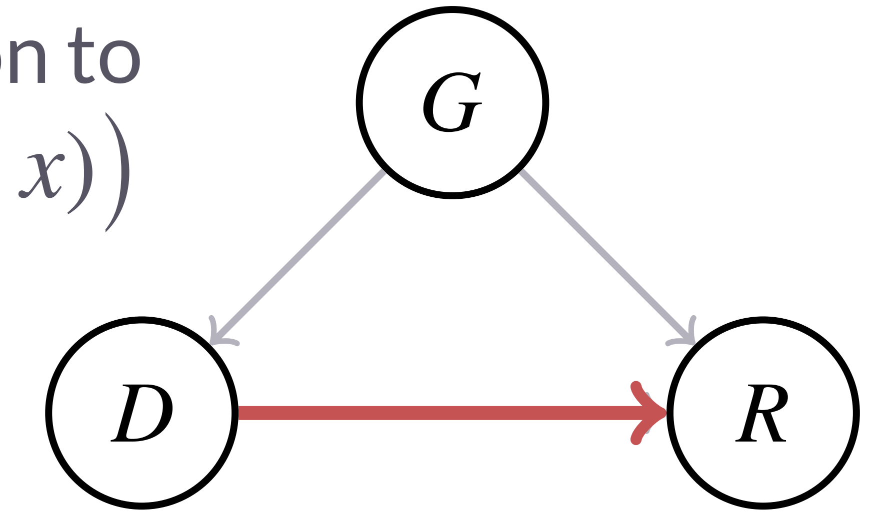
causal model



no contribution to
$P\left(Y = y \mid do(X = x)\right)$

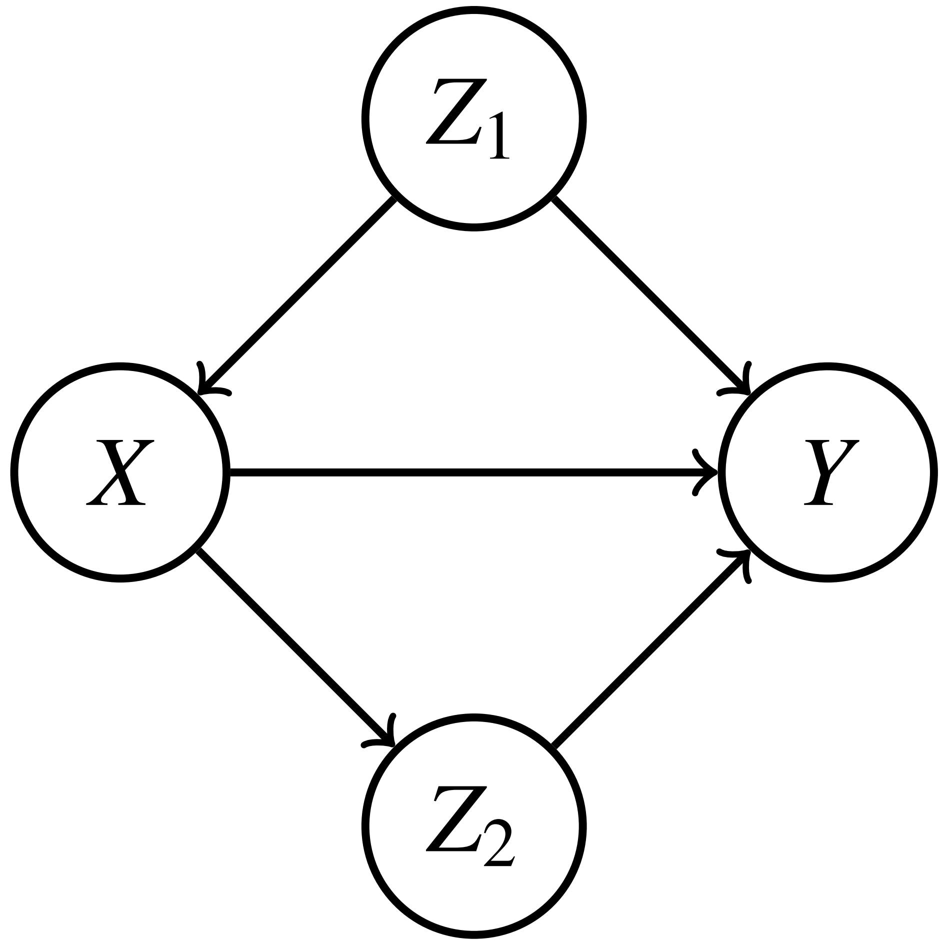


contribution to
$P\left(Y = y \mid do(X = x)\right)$



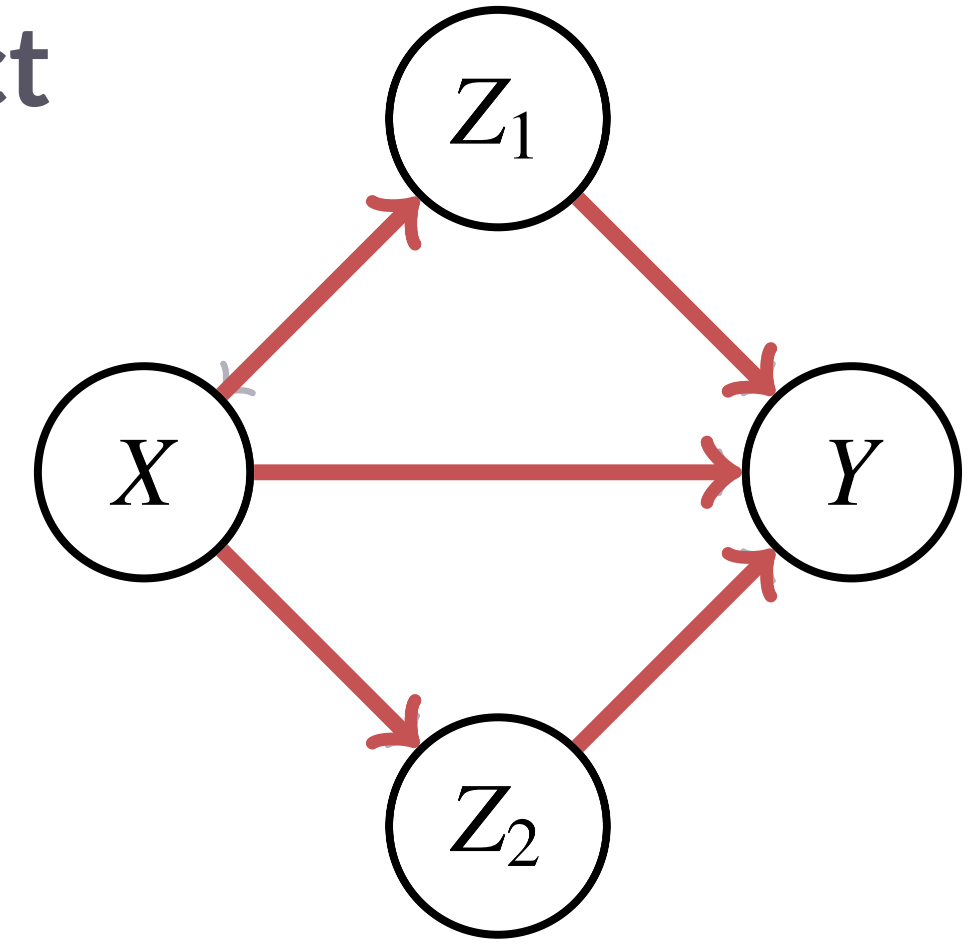
Types of (causal) effects

assumed / true
causal relation



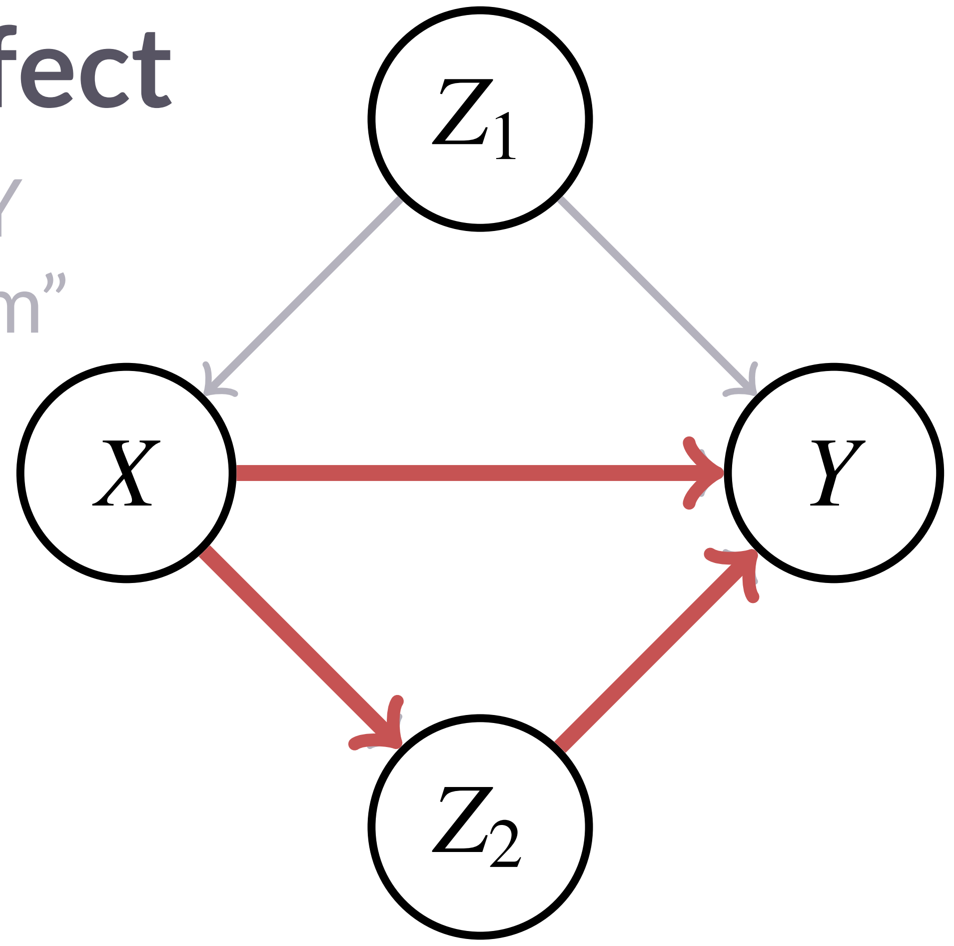
statistical effect

all ways in which
knowing about X
influences (statistical)
predictions about Y



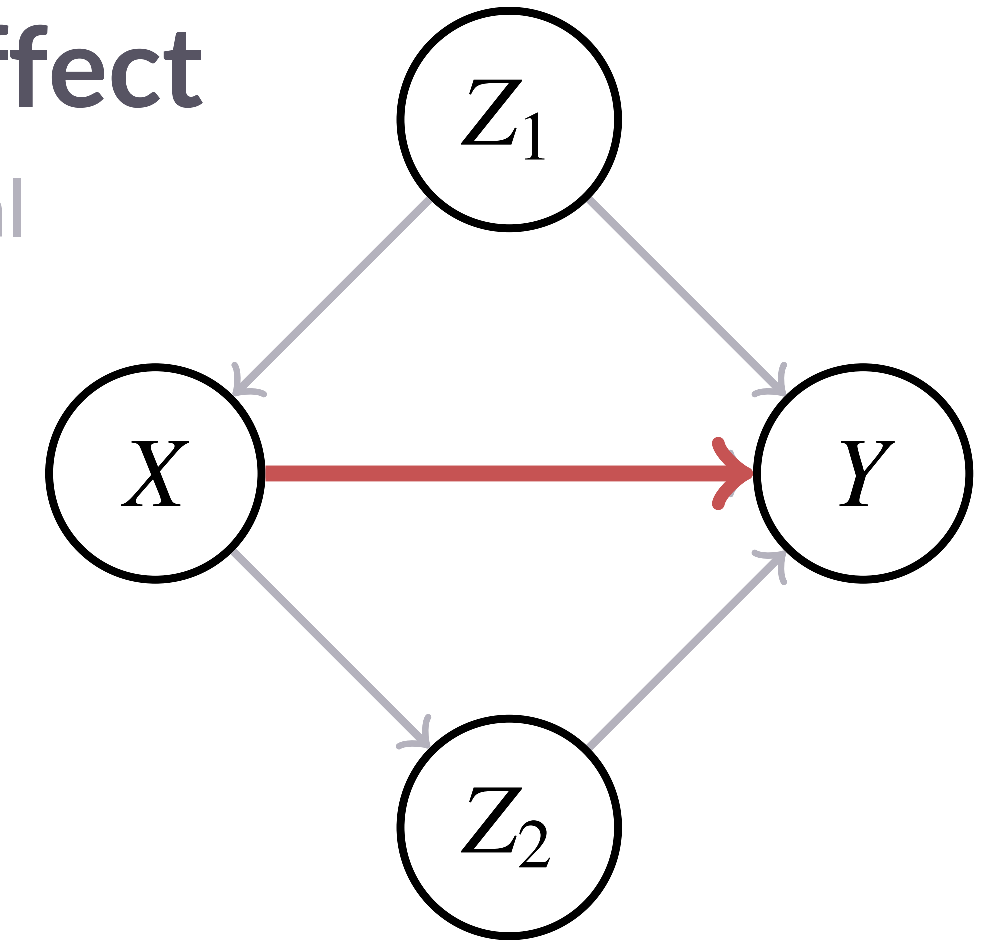
total causal effect

all ways in X affects Y
"causally downstream"



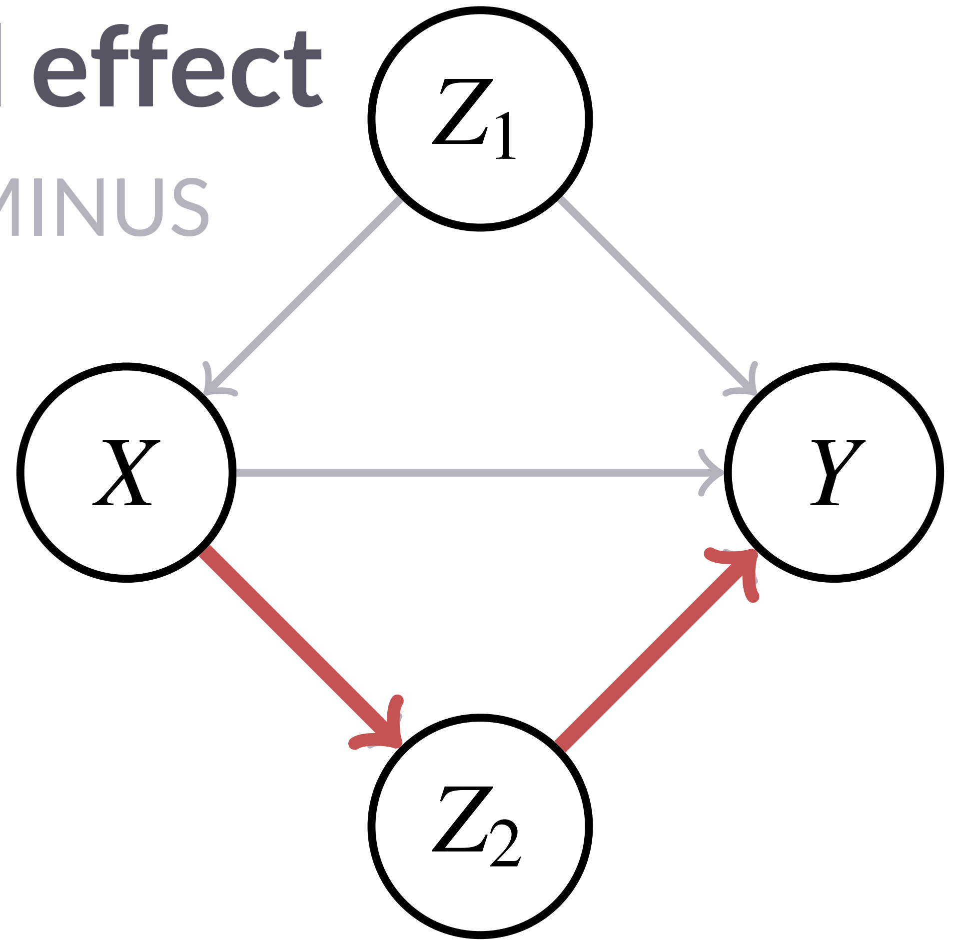
direct causal effect

X's unmediated causal
effect on Y



indirect causal effect

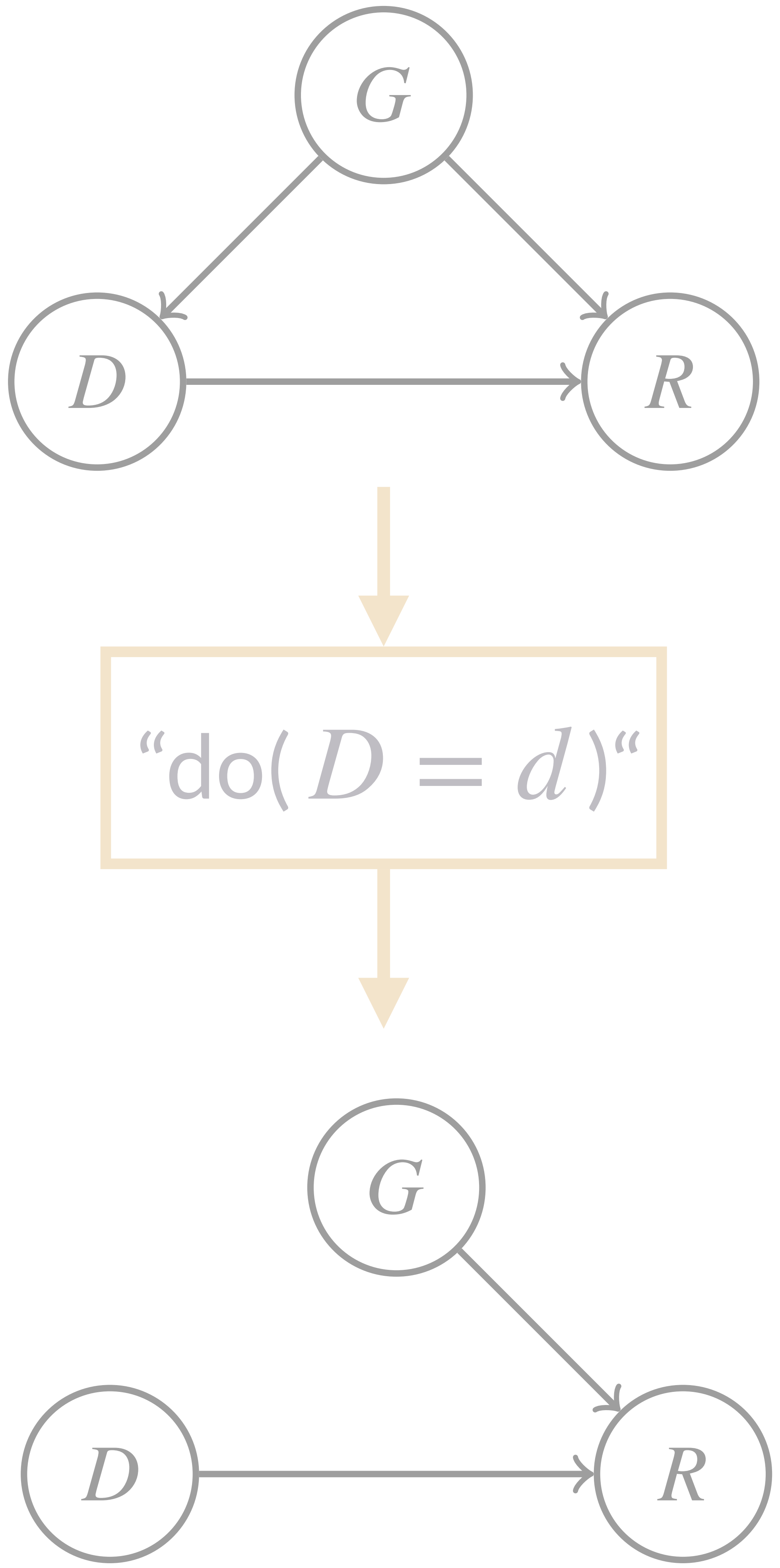
total causal effects MINUS
direct causal effect



causal effects w/ regression modeling

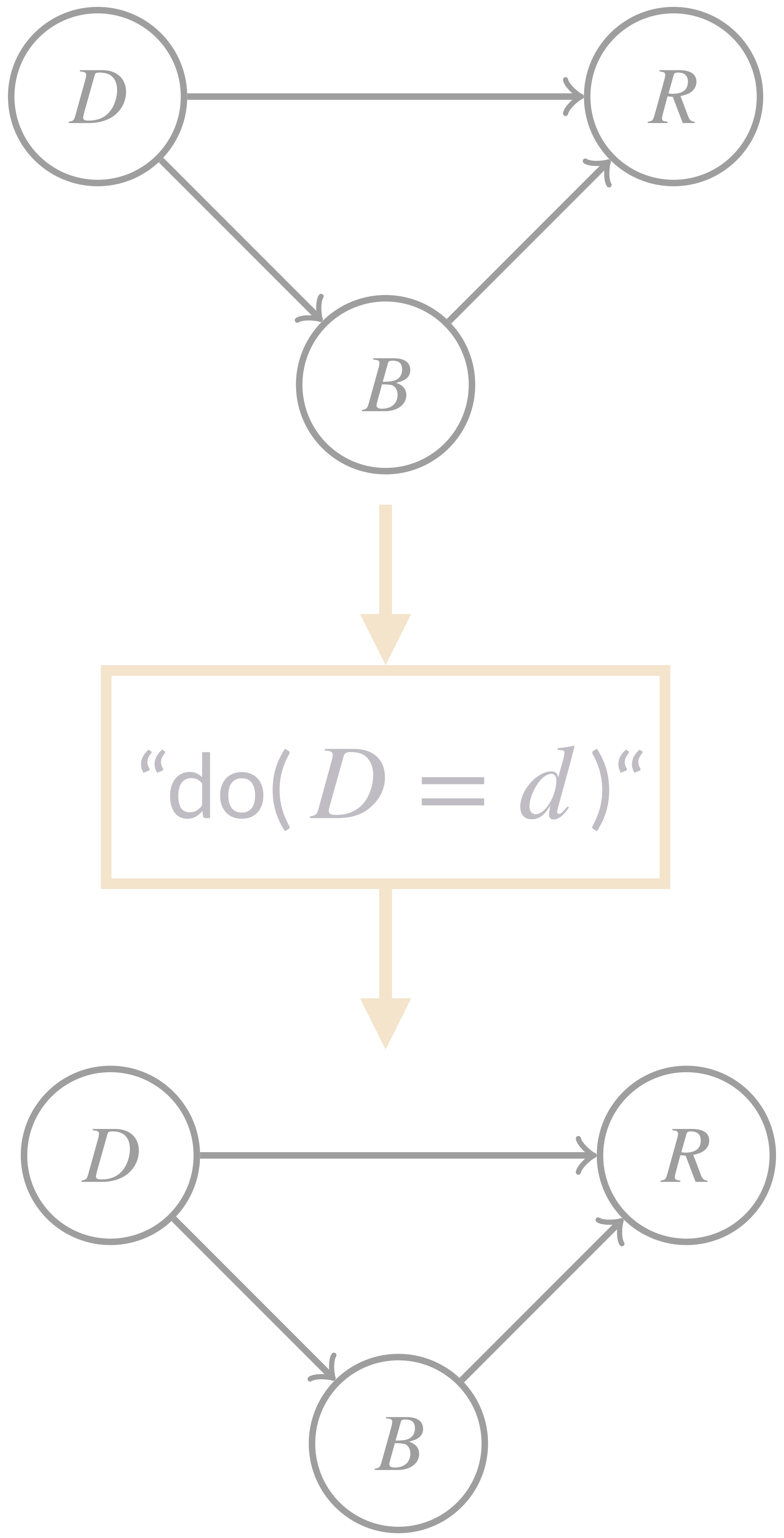
MLE of causal effect

Gender as a confounder



ML-estimate of
causal effect
$0.83 - 0.78 = 0.05$

Blood-pressure as a mediator



ML-estimate of
causal effect
$0.78 - 0.83 = -0.05$

Prepare data

Simpson's paradox

```r
##################################################
# set up the data for SP
##################################################

data_simpsons_paradox <- tibble(
  gender = c("Male", "Male", "Female", "Female"),
  bloodP = c("Low", "Low", "High", "High"),
  drug   = c("Take", "Refuse", "Take", "Refuse"),
  k      = c(81, 234, 192, 55),
  N      = c(87, 270, 263, 80),
  proportion = k/N
)
```

```r
# cast into long format
data_SP_long <- rbind(
  data_simpsons_paradox |> uncount(k) |>
    mutate(recover = TRUE)  |> select(-N, -proportion),
  data_simpsons_paradox |> uncount(N-k) |>
    mutate(recover = FALSE) |> select(-N, -proportion, -k)
)
data_SP_long
```

```
# A tibble: 700 × 4
   gender bloodP drug  recover
   <chr>  <chr>  <chr> <lgl>
 1 Male   Low    Take  TRUE
 2 Male   Low    Take  TRUE
 3 Male   Low    Take  TRUE
 4 Male   Low    Take  TRUE
 5 Male   Low    Take  TRUE
 6 Male   Low    Take  TRUE
 7 Male   Low    Take  TRUE
 8 Male   Low    Take  TRUE
 9 Male   Low    Take  TRUE
10 Male   Low    Take  TRUE
# ℹ 690 more rows
```

Calculating the total causal effect

Case 1: gender as a confound

We want:

$$P(R = 1 \mid do(D = d)) = \sum_{g \in \{0,1\}} P(R = 1 \mid D = d, G = g)\ P(G = g)$$

We do:

1. estimate $P(G)$
   - use intercept-only logistic regression `G ~ 1`
2. estimate $P(R = 1 \mid D = d, G = g)$:
   - use logistic regression model `R ~ D * G`
3. calculate TCE with posterior predictive distributions of these models

Calculating TCE

Case 1: gender as a confound

Step 1: `G ~ 1`

```
niter = 2000

fit_SP_GonIntercept <- brm(
  formula = gender ~ 1,
  data    = data_SP_long,
  family  = bernoulli(link = "logit"),
  iter    = niter
)
```

Step 2: `R ~ D * G`

```
fit_SP_RonGD <- brm(
  formula = recover ~ gender * drug,
  data    = data_SP_long,
  family  = bernoulli(link = "logit"),
  iter    = niter
)
```

Step 3:

sample participants

```
postPred_gender <- tidybayes::predicted_draws(
  object  = fit_SP_GonIntercept,
  newdata = tibble(Intercept = 1),
  value   = "gender",
  ndraws  = niter * 2
  ) |>
  ungroup() |>
  mutate(gender = ifelse(gender, "Male", "Female")) |>
  select(gender)
```

do(drug = 1)

```
# posterior predictive samples for D=1
posterior_DrugTaken <- tidybayes::epred_draws(
  object  = fit_SP_RonGD,
  newdata = postPred_gender |> mutate(drug = "Take"),
  value   = "taken",
  ndraws  = niter * 2
) |> ungroup() |>
  select(taken)
```

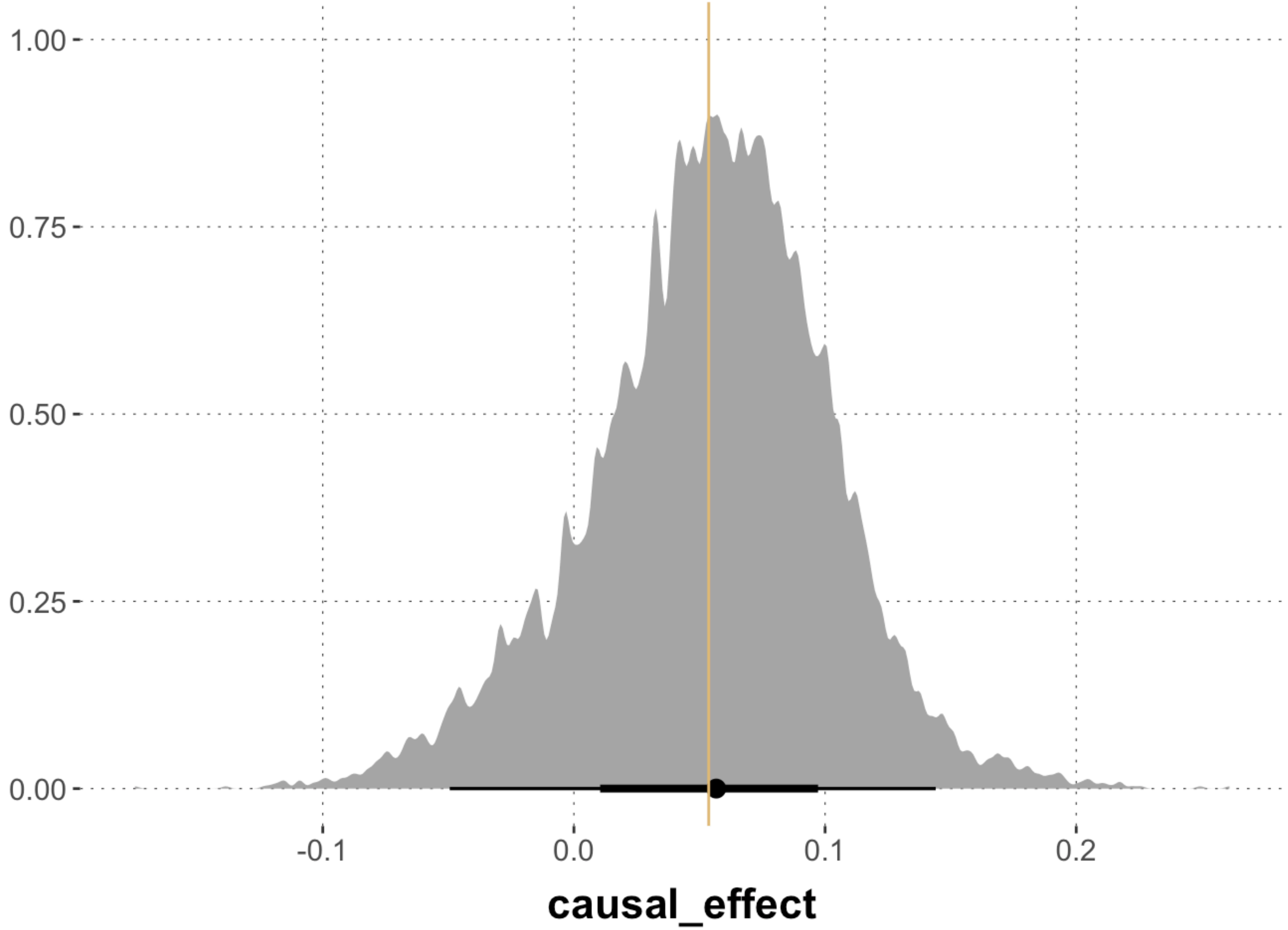
do(drug = 0)

```
# posterior predictive samples for D=0
posterior_DrugRefused <- tidybayes::epred_draws(
  object  = fit_SP_RonGD,
  newdata = postPred_gender |> mutate(drug = "Refuse"),
  value   = "refused",
  ndraws  = niter * 2
) |> ungroup() |>
  select(refused)
```

Calculating TCE

Case 1: gender as a confound

```
# A tibble: 3 × 4
  Parameter     `|95%`  mean `95%|`
  <chr>          <dbl> <dbl>  <dbl>
1 drug_taken    0.687  0.834  0.972
2 drug_refused  0.614  0.780  0.907
3 causal_effect -0.0505 0.0540  0.142
```



ML-estimate of

causal effect

$$0.83 - 0.78 = 0.05$$

Calculating TCE

Case 2: blood-pressure as mediator

We want:

$$P\left(R = r \mid do(D = d)\right) = P\left(R = r \mid D = d\right)$$

We do:

```r
fit_SP_RonBD <- brms::brm(
  formula = recover ~ drug,
  data    = data_SP_long,
  family  = bernoulli(link = "logit"),
  iter    = niter
)
```

```r
posterior_DrugTaken <-
  faintr::extract_cell_draws(fit_SP_RonBD, drug == "Take") |>
  pull(draws) |>
  logistic()

posterior_DrugRefused <-
  faintr::extract_cell_draws(fit_SP_RonBD, drug == "Refuse") |>
  pull(draws) |>
  logistic()

posterior_causalEffect <-
  posterior_DrugTaken - posterior_DrugRefused
```

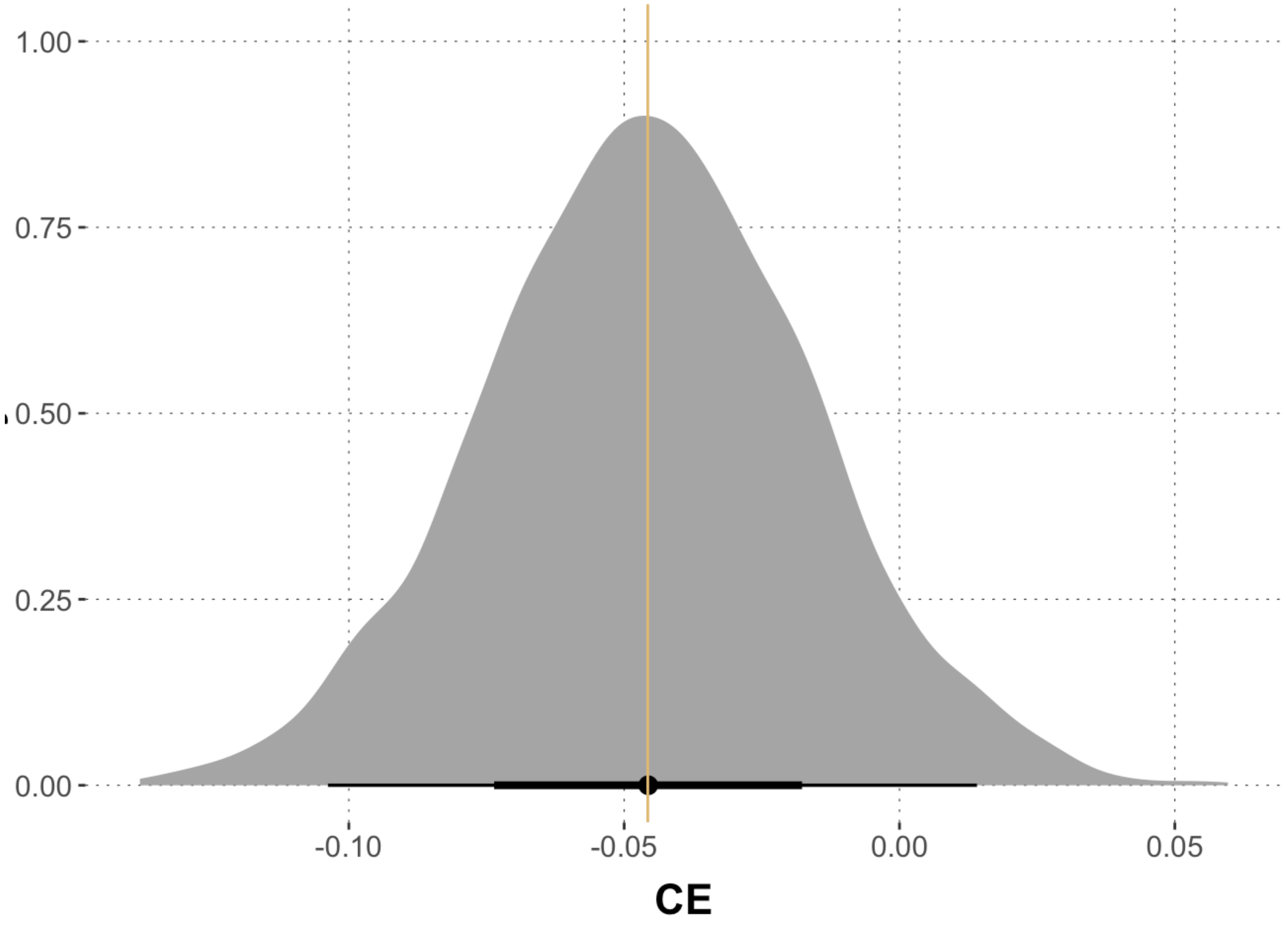
Calculating TCE

Case 2: blood-pressure as mediator

```
# A tibble: 3 × 4
  Parameter     `|95%`    mean `95%|`
  <chr>          <dbl>   <dbl>  <dbl>
1 drug_taken     0.734  0.780  0.822
2 drug_refused   0.786  0.825  0.865
3 causal_effect -0.103 -0.0457 0.0146
```

ML-estimate of
causal effect

$$0.78 - 0.83 = -0.05$$



Causal inference w/ Bayesian regression

summary

- do-calculus tells us when and how we can draw "causal conclusions" from observational data
  - we must specify a causal model
  - readily applicable criteria exist: backdoor, front-door
- uncertainty about causal effect is quantifiable using Bayesian regression modeling
  - but (currently) requires manual labor