



Causal Questions and DAGs in Practice

Causal ML 01 | From a business question to a causal graph: d-separation, the backdoor and frontdoor criteria, minimal adjustment sets, good and bad controls (M-bias, post-treatment variables), testing a DAG against data, DoWhy and dagitty

#CausalML #DAG #Backdoor #Frontdoor #DoWhy

From question to graph

Plumbing plan of a house	Causal DAG
Pipes between rooms	Arrows: direct causal influence
Two taps fed by one tank	Confounder: common cause of T and Y
Closing a valve	Conditioning on a variable that blocks a path
A pipe joint where two flows merge	Collider: opening it lets water flow backwards between sources
Cutting the pipe into one tap and filling it yourself	do(T): intervention removes arrows into T

WHERE THE ANALOGY BREAKS A plumber can open walls to check pipes; most causal arrows cannot be seen, only argued from domain knowledge and partly tested against data.

A DAG is a wiring diagram you commit to before looking at effects

Before Adjustment sets from a DAG, can you explain...

	Not yet	Partly	Yes
MUST KNOW Confounder, mediator, collider roles Learn: machine-learning 16	☐	☐	☐
MUST KNOW Conditional independence Learn: maths 08	☐	☐	☐
MUST KNOW Linear regression coefficients Learn: machine-learning 02	☐	☐	☐
HELPFUL Bayesian networks as factorisations Learn: symbolic-and-probabilistic-ai	☐	☐	☐

Check before reading; recalled in one line where used

- Unit and time:** who is treated and when (user-week, store-month); the graph's nodes are measured at those times.
- Treatment as an intervention you could perform:** 'send the e-mail' is a treatment; 'being loyal' is not.
- Outcome and horizon:** purchase within 14 days, not 'engagement'.
- Estimand:** total effect (all paths from T) or direct effect (paths not through a mediator)? Sheet 02 names them.
- Variables available at decision time:** anything measured after T is a candidate bad control.

Writing the causal question before the graph

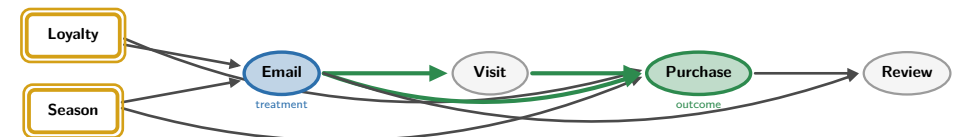
Triple	Shape	Unconditioned	Conditioned on middle
Chain	$A \rightarrow M \rightarrow B$	open (A and B associated)	blocked
Fork	$A \leftarrow C \rightarrow B$	open	blocked
Collider	$A \rightarrow K \leftarrow B$	blocked	open (also if a descendant of K is conditioned on)

d-separation in three patterns: every path is a chain of these triples

Rule

A path is blocked by a set Z if at least one of its triples is blocked. T and Y are d-separated by Z if every path is blocked. The backdoor criterion asks this only for paths starting $T \leftarrow$.

Backdoor adjustment on a real graph



■ treatment ■ outcome ■ descendant of treatment ■ collider / adjusted (box)

kind	path	colliders	status given $Z = \{\text{Loyalty, Season}\}$
backdoor	$\text{Email} \leftarrow \text{Loyalty} \rightarrow \text{Purchase}$	---	blocked: Loyalty adjusted
backdoor	$\text{Email} \leftarrow \text{Season} \rightarrow \text{Purchase}$	---	blocked: Season adjusted
causal	$\text{Email} \rightarrow \text{Visit} \rightarrow \text{Purchase}$	---	open (keep: the effect)
causal	$\text{Email} \rightarrow \text{Purchase}$	---	open (keep: the effect)

$Z = \{\text{Loyalty, Season}\}$ satisfies the backdoor criterion for $\text{Email} \rightarrow \text{Purchase}$.

$P(y \mid \text{do}(t)) = \sum \text{over loyalty, season of } P(y \mid t, \text{loyalty, season}) \cdot P(\text{loyalty, season})$ (backdoor adjustment)

A path is blocked by Z when it holds a non-collider in Z, or a collider that is not in Z and has no descendant in Z (d-separation).

where T, Y = treatment (Email) and outcome (Purchase); Z = adjustment set (no descendant of T allowed); backdoor path = path from T whose first arrow points into T; collider = node with both path arrows pointing in ($\rightarrow c \leftarrow$); do(t) = intervention setting $T = t$

Does a promotional e-mail cause purchases? Two backdoor paths (via Loyalty, via Season) are blocked by {Loyalty, Season}. Visit is a mediator, Review a collider: adjusting for either biases the total effect

Criterion	Formula	When it applies
Backdoor	$P(y do(t)) = \sum_z P(y t, z) P(z)$	Z blocks all backdoor paths and contains no descendant of T
Frontdoor	$P(y do(t)) = \sum_m P(m t) \sum_{t'} P(y m, t') P(t')$	M intercepts all directed T → Y paths; T → M unconfounded; M → Y backdoor blocked by T

where t, t' = treatment values (t' sums over all of them) (0/1); y = outcome value (-); z = values of the adjustment set Z (-); m = values of the mediator M (-); $P(\cdot | do(t))$ = distribution if T were set to t for everyone (probability)

The two identification formulas of this sheet

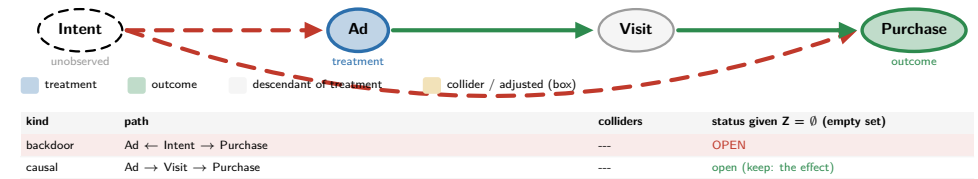
Confounder $T \leftarrow C \rightarrow Y$: adjust Tell: Does the variable sit before T or after T on the path?	vs	Mediator $T \rightarrow M \leftarrow Y$: do not adjust for the total effect
Pre-treatment confounder common cause: adjust Tell: Is it caused by two separate causes of T and Y?	vs	M-bias collider $A \rightarrow M \leftarrow B$ with $A \rightarrow T, B \rightarrow Y$: adjusting opens a path
Outcome predictor $Z \rightarrow Y$ only: adjusting is neutral and cuts variance Tell: Does the variable point into Y or into T?	vs	Treatment predictor (instrument) $Z \rightarrow T$ only: adjusting adds variance and amplifies residual bias
Proxy of a confounder partially blocks the path: usually adjust Tell: Was it measured before T happened?	vs	Descendant of the outcome $Y \rightarrow D$: conditioning selects on Y, a collider-like bias

Good and bad controls (after Cinelli, Forney and Pearl, 2022)

⚠ Table 2 fallacy

One regression with T and its adjustment set Z gives T's effect. The coefficients of Z are **not** causal effects of Z (each needs its own adjustment set). Never print all coefficients of a confounder-adjusted model as 'drivers'.

🔍 Frontdoor: when the confounder is hidden



No valid backdoor adjustment set among the observed nodes: an unobserved confounder keeps a backdoor path open (Ad → Purchase not identified this way).

- open: Ad ← Intent → Purchase (no adjusted non-collider)

$P(y | do(t)) \neq P(y | t, z)$ here: the set does not block every backdoor path

A path is blocked by Z when it holds a non-collider in Z, or a collider that is not in Z and has no descendant in Z (d-separation).

where T, Y = treatment (Ad) and outcome (Purchase); Z = adjustment set (no descendant of T allowed); backdoor path = path from T whose first arrow points into T; collider = node with both path arrows pointing in ($\rightarrow c \leftarrow$); $do(t)$ = intervention setting T = t

Ad exposure → site visit → purchase, with unobserved purchase intent driving both ad targeting and purchases. No observed set blocks the backdoor via Intent; the frontdoor path through Visit still identifies the effect

PROBLEM Ad X, visit M, purchase Y (all 0/1); intent is unobserved. Effect of the ad on purchase probability?

- GIVEN**
- $P(X = 1) = 0.5$
 - $P(M = 1 | X = 1) = 0.8, P(M = 1 | X = 0) = 0.2$
 - $P(Y = 1 | M, X): (1,1) 0.30; (1,0) 0.20; (0,1) 0.10; (0,0) 0.05$

1 Inner sum for M = 1

$$0.5 \times 0.30 + 0.5 \times 0.20 = 0.25$$

Why: $P(Y | M = 1)$ averaged over X, which blocks M's backdoor via Intent

2 Inner sum for M = 0

$$0.5 \times 0.10 + 0.5 \times 0.05 = 0.075$$

Why: same for M = 0

3 $P(Y = 1 | do(X = 1))$

$$0.8 \times 0.25 + 0.2 \times 0.075 = 0.215$$

Why: weight by how the ad moves M

4 $P(Y = 1 | do(X = 0))$

$$0.2 \times 0.25 + 0.8 \times 0.075 = 0.110$$

Why: same with no ad

5 Naive comparison

$$P(Y|X=1) = 0.26, P(Y|X=0) = 0.08 \rightarrow 0.18$$

Why: mixes the effect with intent-driven targeting

ANSWER Causal effect = $0.215 - 0.110 = +10.5$ points (the naive 18 points overstate it by 7.5)

CHECK ✓ Valid only if the ad acts on purchases solely through visits and nothing confounds ad → visit

Frontdoor estimate by hand with binary variables

🔗 Testing and coding the graph

Check	How	What it proves
Implied independences	dagitty impliedConditional Independencies, DoWhy graph refuter	a rejected one falsifies the DAG
Placebo treatment	replace T by a random permutation	estimate should drop to about 0
Negative-control outcome	an outcome T cannot affect (last year's spend)	non-zero effect = residual confounding
Random common cause	add a simulated confounder	a stable estimate is a weak robustness sign
E-value (VanderWeele and Ding, 2017)	$E = RR + \sqrt{(RR \times (RR - 1))}$, $RR = \text{risk ratio}$	how strong a hidden confounder must be to explain the effect away
Missing arrows	cannot be checked if no independence is implied	domain knowledge only

What you can and cannot check against data

```
DoWhy (PyWhy, API as of 2026): model, identify, estimate, refute
from dowhy import CausalModel
```

```
g = """digraph { Loyalty -> Email; Loyalty -> Purchase;
  Season -> Email; Season -> Purchase;
  Email -> Visit; Visit -> Purchase;
  Email -> Purchase; }"""
m = CausalModel(data=df, treatment="Email",
  outcome="Purchase", graph=g)
estimand = m.identify_effect() # backdoor set
print(estimand) # lists Loyalty, Season
est = m.estimate_effect(
  estimand,
  method_name="backdoor.linear_regression")
ref = m.refute_estimate(
  estimand, est,
  method_name="placebo_treatment_refuter")
```

⊕ **Your manager asks to 'control for everything' in the churn model. What do you answer?**

Adjust only for a valid set: pre-treatment confounders. Variables measured after the treatment (usage after the discount, support tickets) are mediators or colliders and bias the effect. I would draw the DAG with the domain team, read the minimal adjustment set with DoWhy or dagitty, and report a refutation and an E-value.

🚩 Myths and recap

YOU MAY HEAR

"More covariates always reduce bias"

BUSTED

✓ ACTUALLY

Colliders, mediators and M-bias variables add bias

Pick an adjustment set from the graph, not from feature importance.

YOU MAY HEAR

"Pre-treatment variables are always safe"

BUSTED

✓ ACTUALLY

A pre-treatment collider (M-bias) is not

In practice M-bias is often small, but instruments and colliders should still be left out.

YOU MAY HEAR

"The DAG is validated if the effect looks plausible"

BUSTED

✓ ACTUALLY

Only implied independences and refuters test it

Plausibility is not evidence.

YOU MAY HEAR

"A hidden confounder makes the effect unidentifiable"

BUSTED

✓ ACTUALLY

Not always: frontdoor and instruments can still identify it

At the cost of other strong assumptions.

Tick a box only when you could explain the point to someone else



Write unit, treatment, outcome, horizon, estimand first

then draw the graph



Adjust for a set that blocks every backdoor path

and contains no descendant of T



Leave out mediators, colliders, instruments

and anything measured after T



Test the graph's implied independences

a rejection means fix the graph



Refute and bound

placebo, negative controls, E-value

Q1 When is a path blocked by Z?

A When it contains a chain or fork whose middle is in Z, or a collider not in Z with no descendant in Z.

Q2 Backdoor criterion in one line?

A Z blocks every path $T \leftarrow \dots \rightarrow Y$ and contains no descendant of T.

Q3 Why is Review a bad control in the e-mail DAG?

A It is a collider ($\text{Email} \rightarrow \text{Review} \leftarrow \text{Purchase}$): conditioning on it opens a spurious Email--Purchase path.

Q4 Frontdoor answer for the ad example?

A +10.5 points: 0.215 - 0.110, against a naive 18 points.

Q5 What is M-bias?

A Bias from adjusting for a pre-treatment collider whose parents cause T and Y separately.

Q6 Why not adjust for an instrument?

A It explains T only: variance rises and any residual confounding bias is amplified.

Q7 What is the Table 2 fallacy?

A Reading the confounders' coefficients of the treatment model as causal effects.

Q8 E-value for a risk ratio of 2?

A $2 + \sqrt{2 \times 1} \approx 3.41$: a hidden confounder would need $RR \geq 3.41$ with both T and Y to explain the effect away.

Q9 What does $do(T = t)$ do to the graph?

A Deletes every arrow into T and sets T to t.