

# Causality: структурные модели причинно-следственных связей

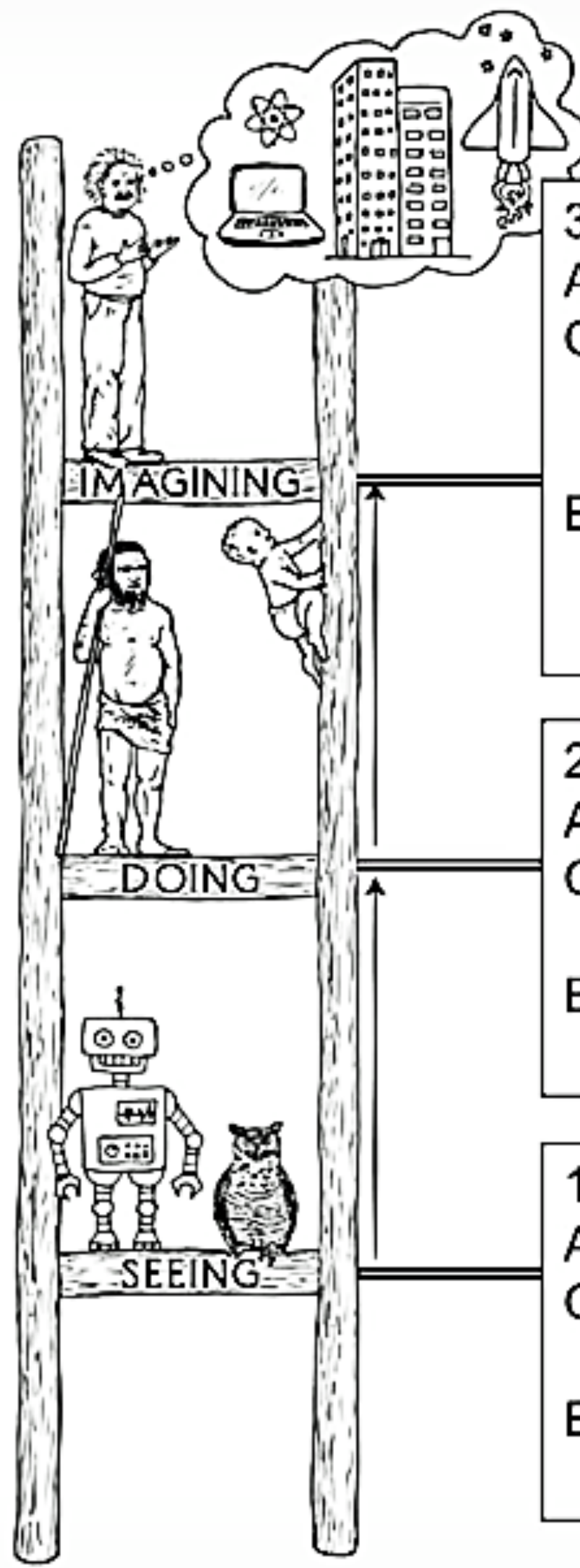
БПМИ192, НИС

Мария Тимонина

Юлия Кокорина

Дмитрий Поздеев

# 3-LEVEL HIERARCHY



## 3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done . . . ? Why?*

(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache?  
Would Kennedy be alive if Oswald had not killed him? What if I had not smoked the last 2 years?

## 2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do . . . ? How?*

(What would Y be if I do X?)

EXAMPLES: If I take aspirin, will my headache be cured?  
What if we ban cigarettes?

## 1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see . . . ?*

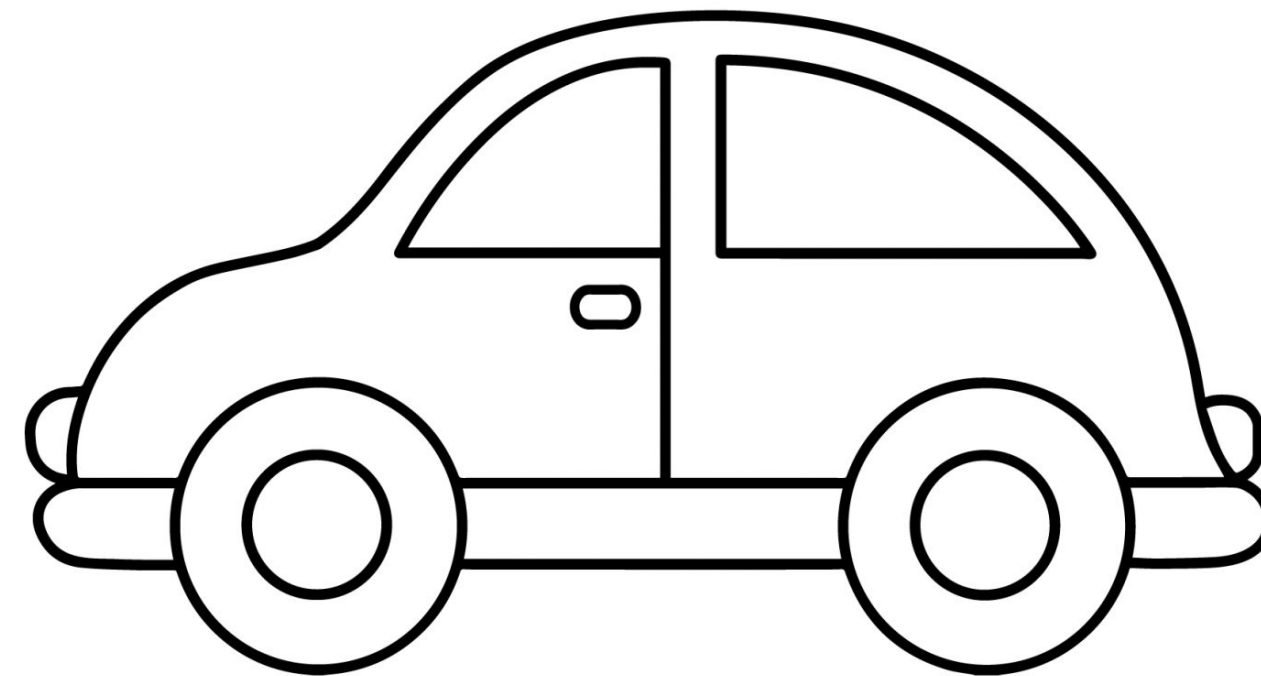
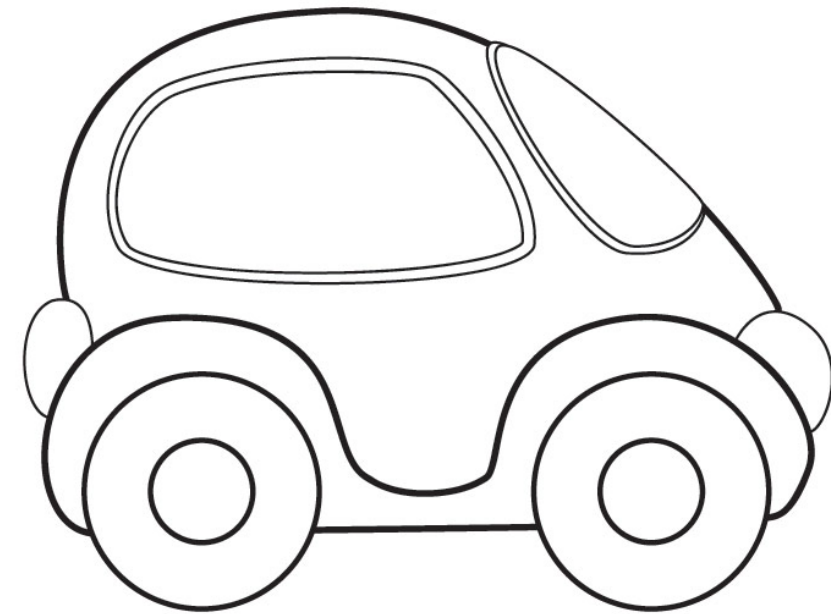
(How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?  
What does a survey tell us about the election results?

The Book of Why: The New Science of Cause and Effect ·  
Judea Pearl, Dana Mackenzie · 2018

# Наблюдаемое VS возможное гипотетически

- Попадают ли 16-летние водители в ДТП чаще, чем 18-летние?



$$P(X|Y = 16) - P(X|Y = 18)$$

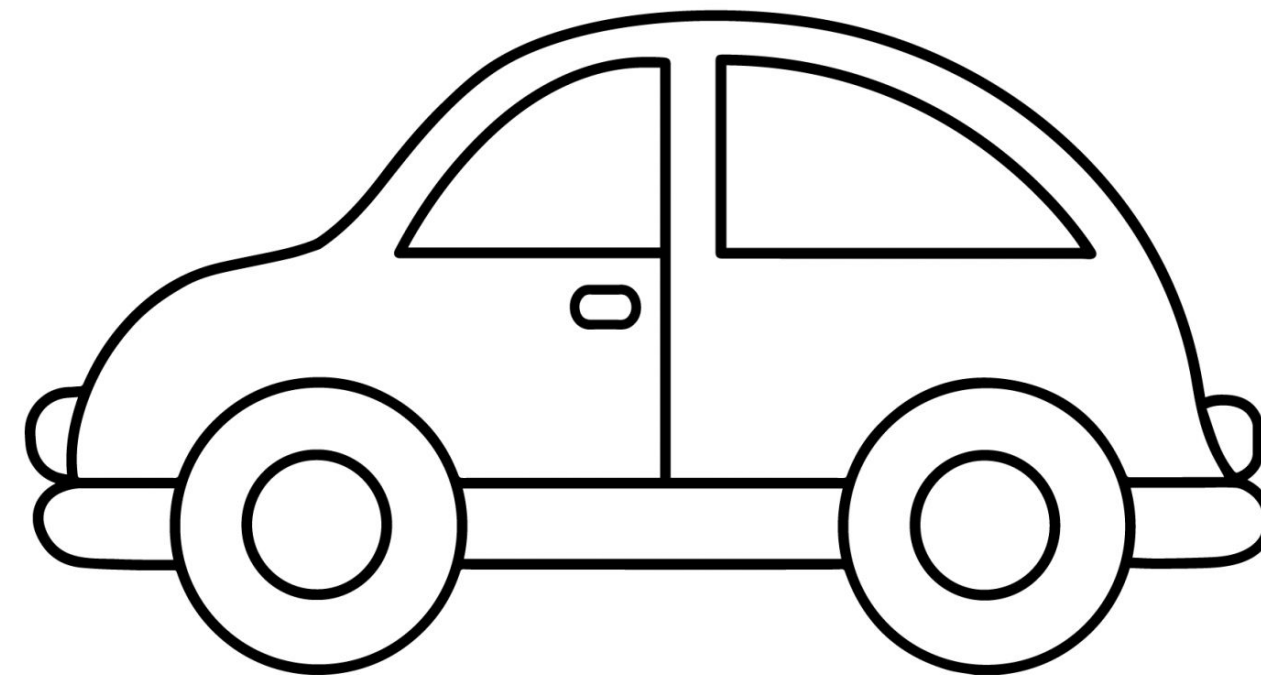
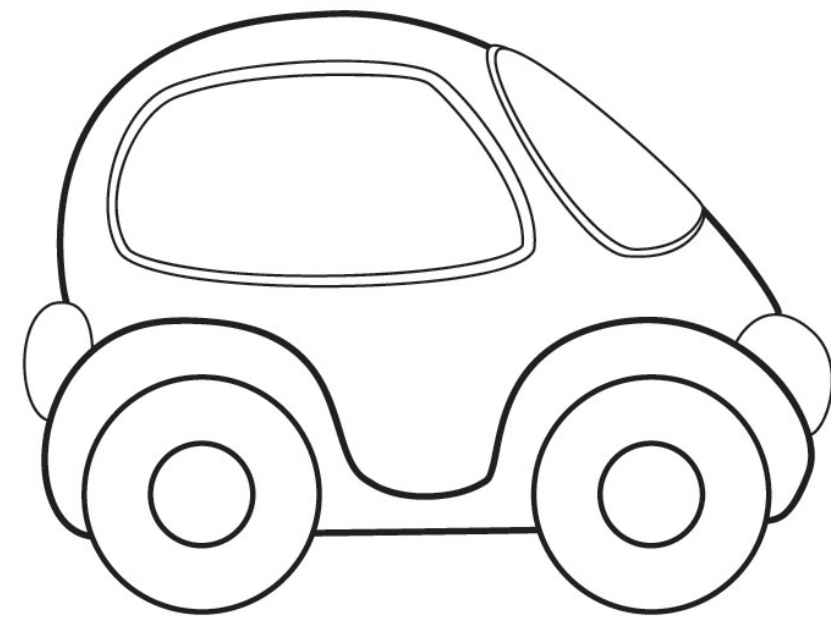
**выборочные условные вероятности**



# Наблюдаемое VS возможное гипотетически

- Попадают ли 16-летние водители в ДТП чаще, чем 18-летние?

$$P(X|Y = 16) - P(X|Y = 18)$$



Вопросы бывают устроены сложнее:

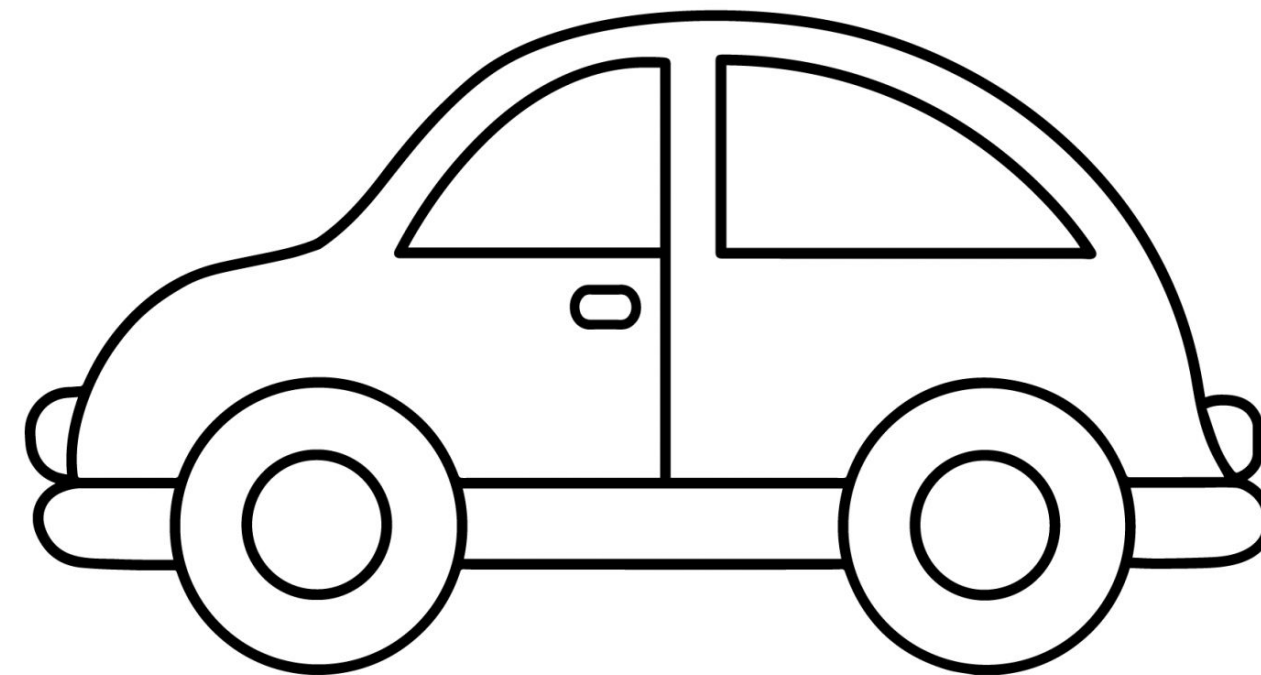
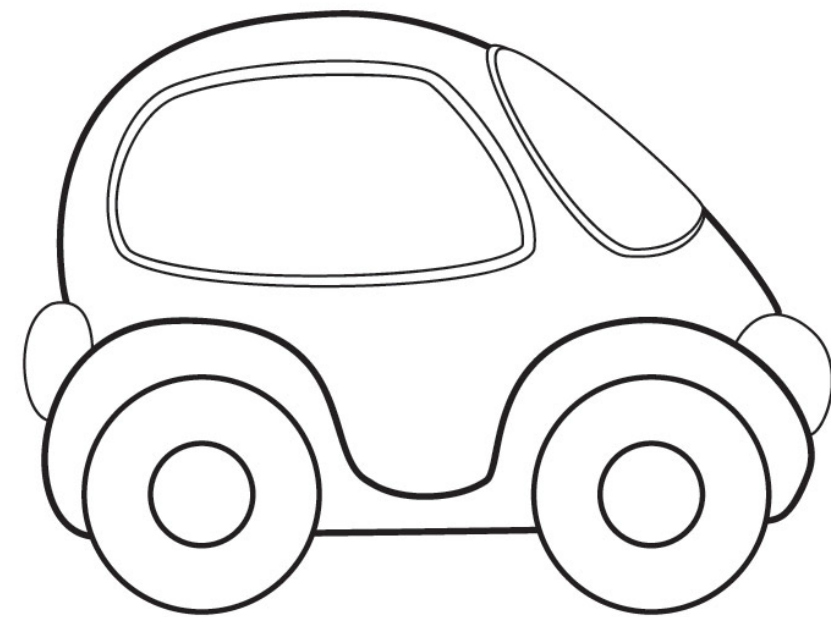
Снизится ли количество фиксируемых ежегодно нарушений на дорогах, если повысить минимальный возраст получения прав?

**Intervention – возможное действие, вмешательство**

# Наблюдаемое VS возможное гипотетически

- Попадают ли 16-летние водители в ДТП чаще, чем 18-летние?

$$P(X|Y = 16) - P(X|Y = 18)$$



Вопросы бывают устроены сложнее:

Снизится ли количество фиксируемых ежегодно нарушений на дорогах, если повысить минимальный возраст получения прав?

**Intervention – возможное действие, вмешательство**

Было ли бы число аварий на дорогах меньшим сейчас, если бы такой закон ввели три года назад?

# Structural Causal model

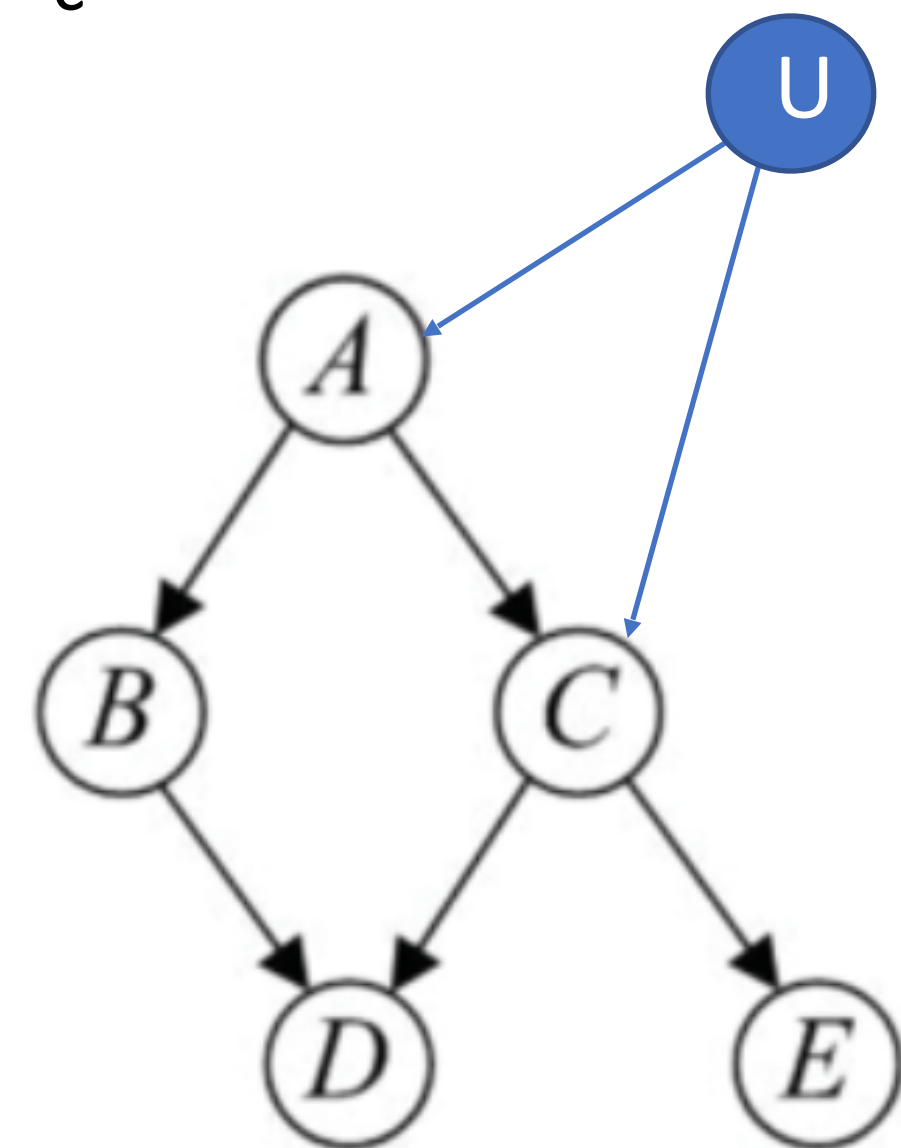
- ациклический граф
- последовательность инструкций для генерации совместного распределения, начиная с независимых шумовых переменных

$$X_i := f_i(P_i, U_i), \quad i = 1, \dots, d.$$

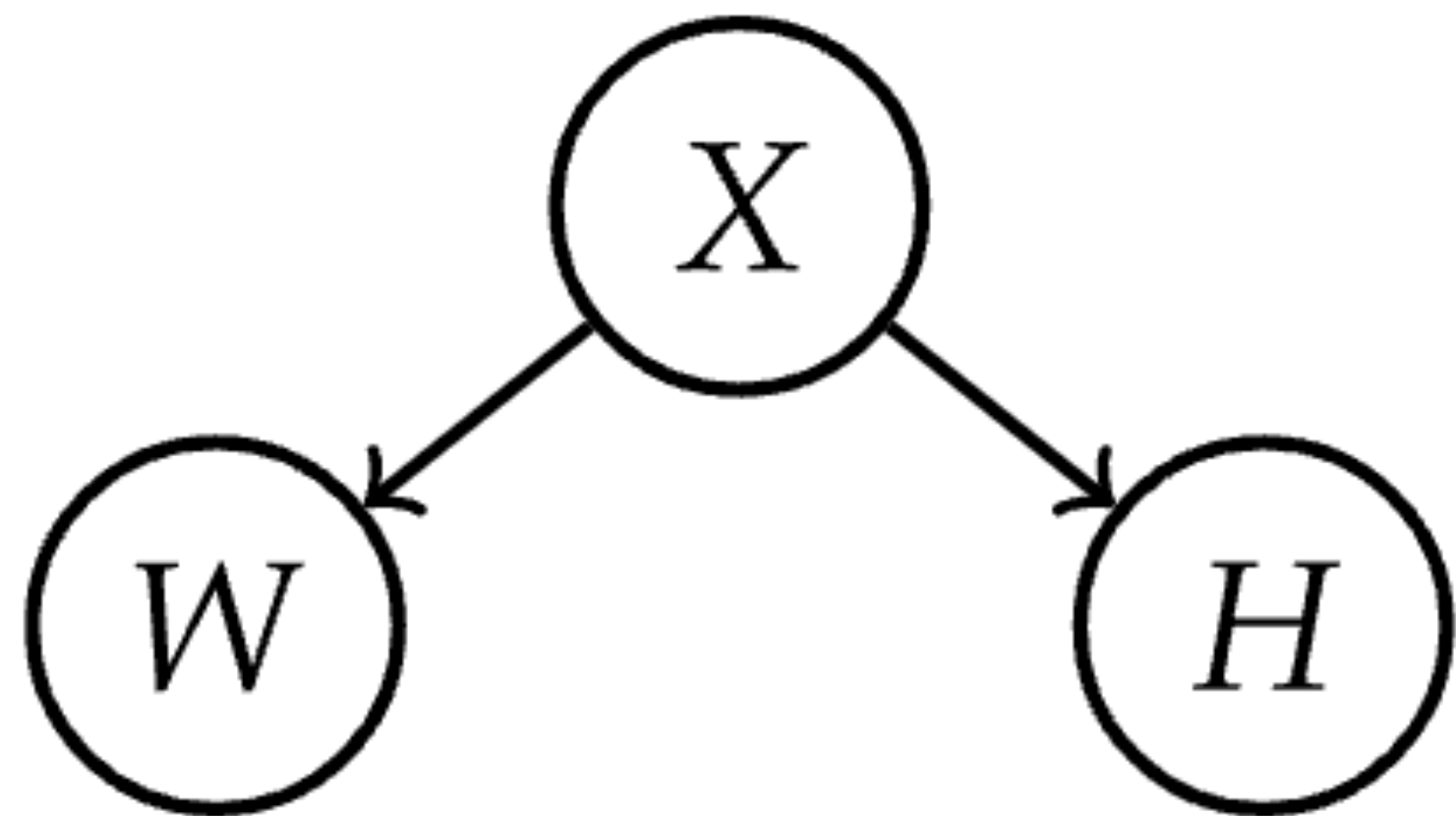
где

$P_i \subseteq \{X_1, \dots, X_d\}$  - это подмножество родительских вершин  $i$ -ой вершины графа, «непосредственных причин»

$U_1, \dots, U_d$  - это шум, внешние случайные величины. От них мы потребуем совместной независимости.



# Разберем модельную задачу



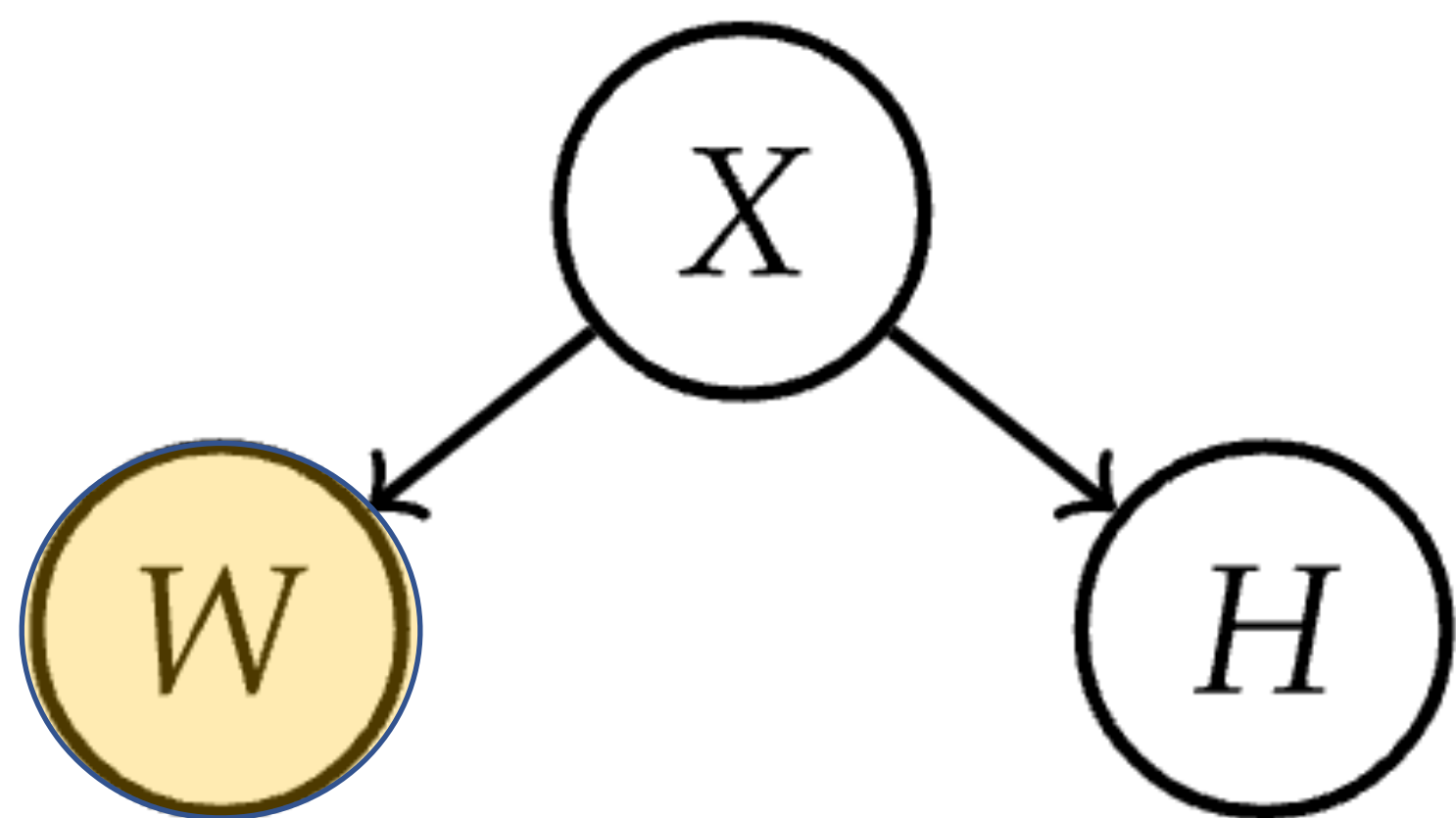
$$X := U_1 \sim B(1/2)$$

$$W := \text{if } X = 1 \text{ then } 0 \text{ else } U_2 \sim B(1/3)$$

$$H := \text{if } X = 1 \text{ then } 0 \text{ else } U_3 \sim B(1/3)$$

Интуитивно переменные  $W$  и  $H$  – наличие у человека лишнего веса и проблем с сердцем – не должны коррелировать

# Разберем модельную задачу



$$W = 1$$

$$X := U_1 \sim \text{B}(1/2)$$

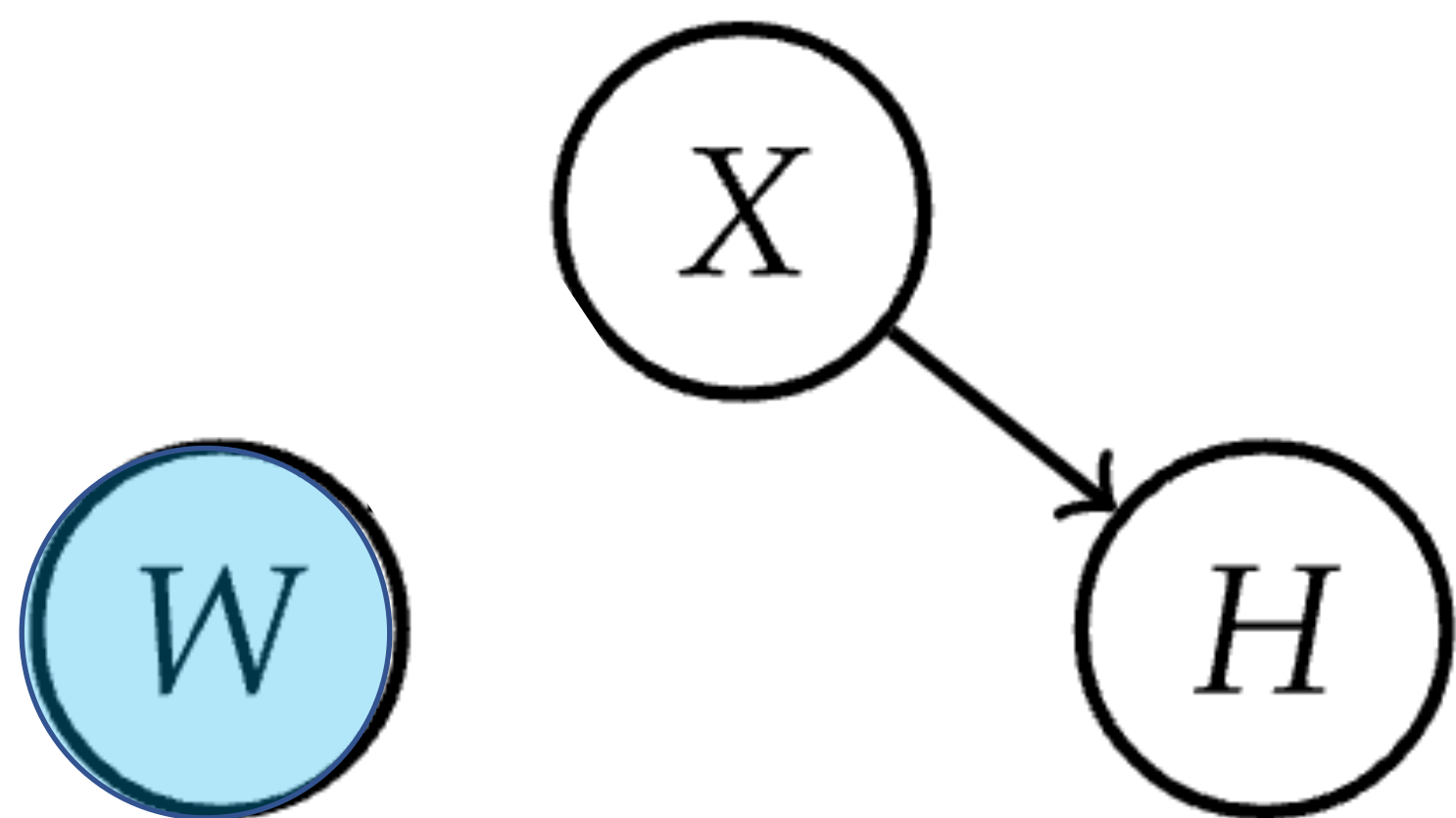
$$W := \text{if } X = 1 \text{ then } 0 \text{ else } U_2 \sim \text{B}(1/3)$$

$$H := \text{if } X = 1 \text{ then } 0 \text{ else } U_3 \sim \text{B}(1/3)$$

$$\mathbb{P}(H|W = 1) =$$



# Разберем модельную задачу



$$W := 1$$

Распределение изменилось!

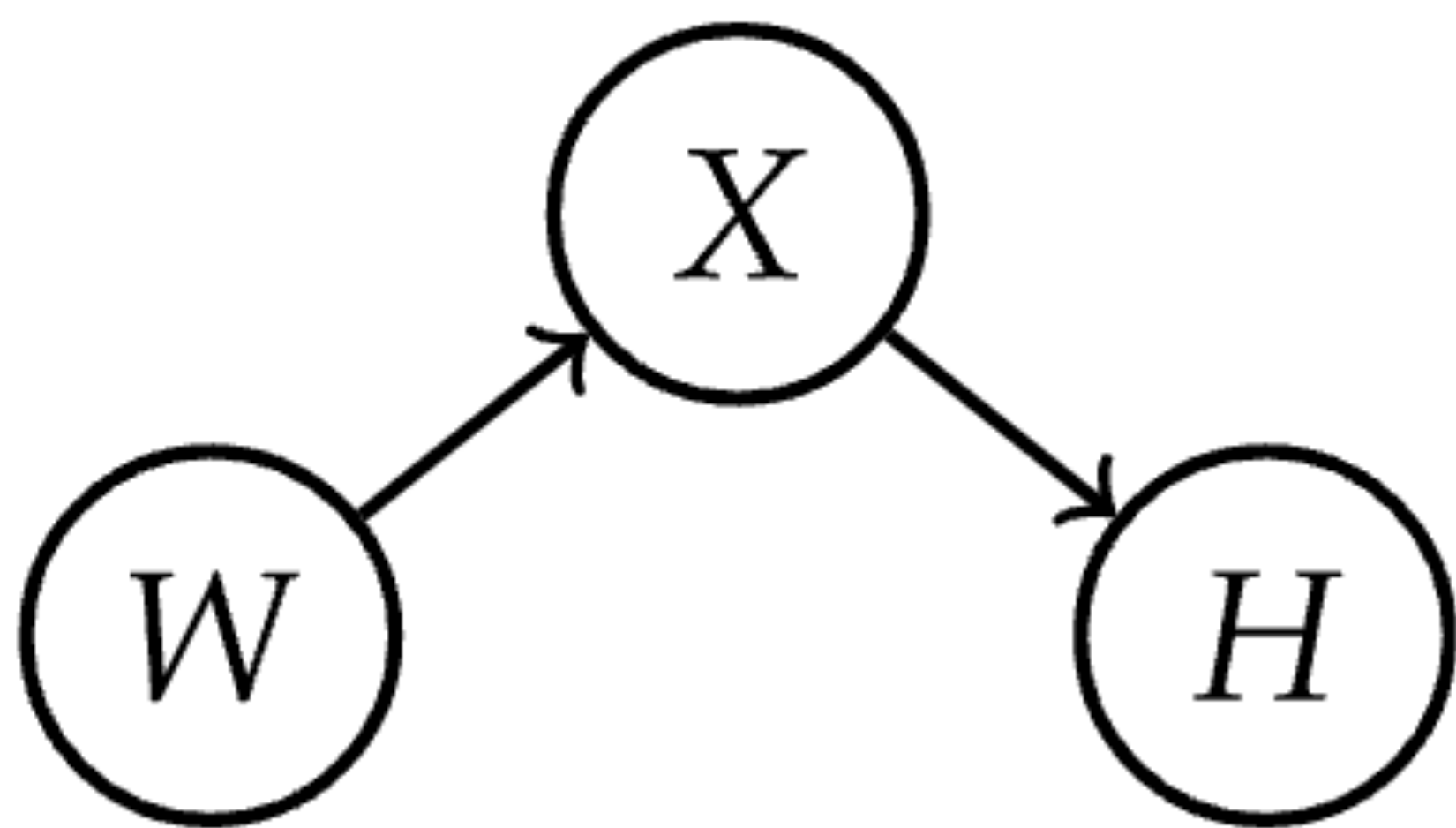
$$X := U_1 \sim \text{B}(1/2)$$

$$W := 1$$

$$H := \text{if } X = 1 \text{ then } 0 \text{ else } U_3 \sim \text{B}(1/3)$$

$$\mathbb{P}(H | do(W := 1)) =$$

## Другой тип зависимости



$$W = 1$$

$$W := 1$$

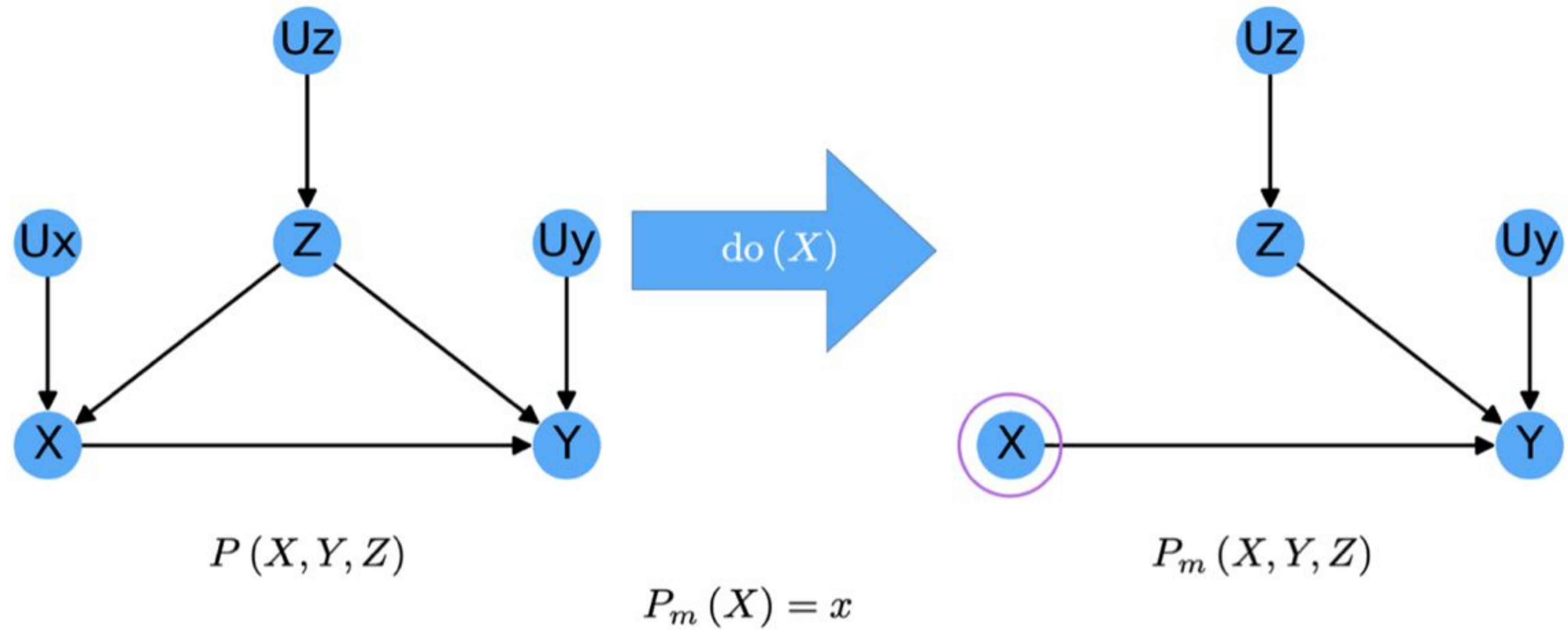
$$U_1 \sim B(1/2), U_2 \sim B(1/3), U_3 \sim B(1/3)$$

$$2. W := U_2$$

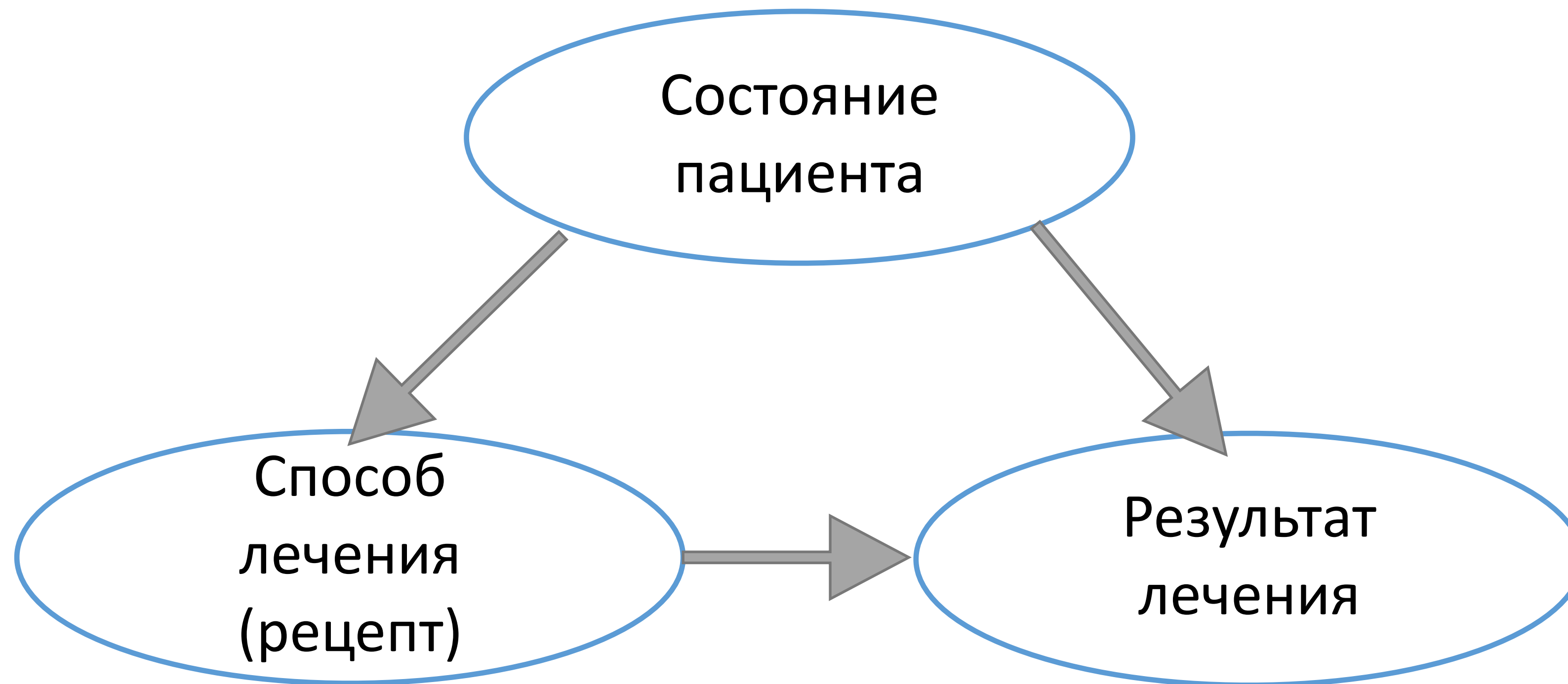
$$3. X := \text{if } W = 0 \text{ then } 0 \text{ else } U_1$$

$$4. H := \text{if } X = 1 \text{ then } 0 \text{ else } U_3$$

# Do-operator

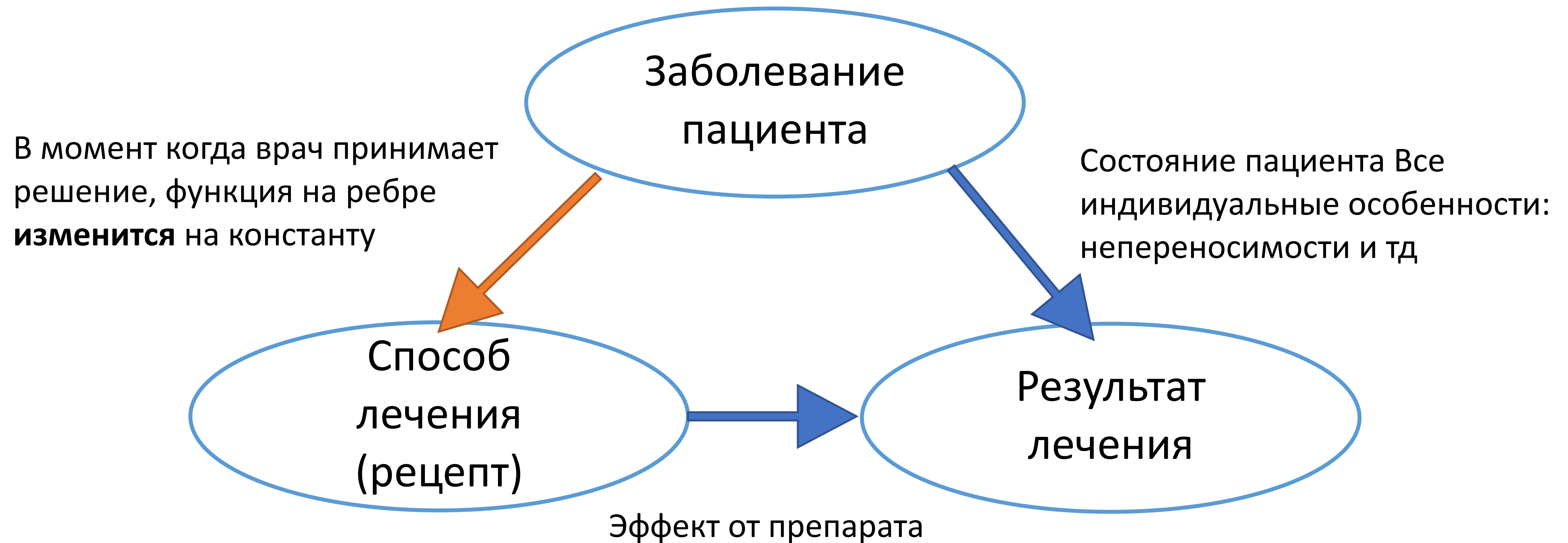


# Не одинаковые связи





# Confounder – связанные переменные



# Quiz: Будем считать $P(y | do(x))$ или $P(y | x)$ ?

- Кейс 1: Доктор

Вы работаете в больнице. Ваш пациент сдает анализ, по результатам которого Вы хотите выбрать лучший способ его лечения.

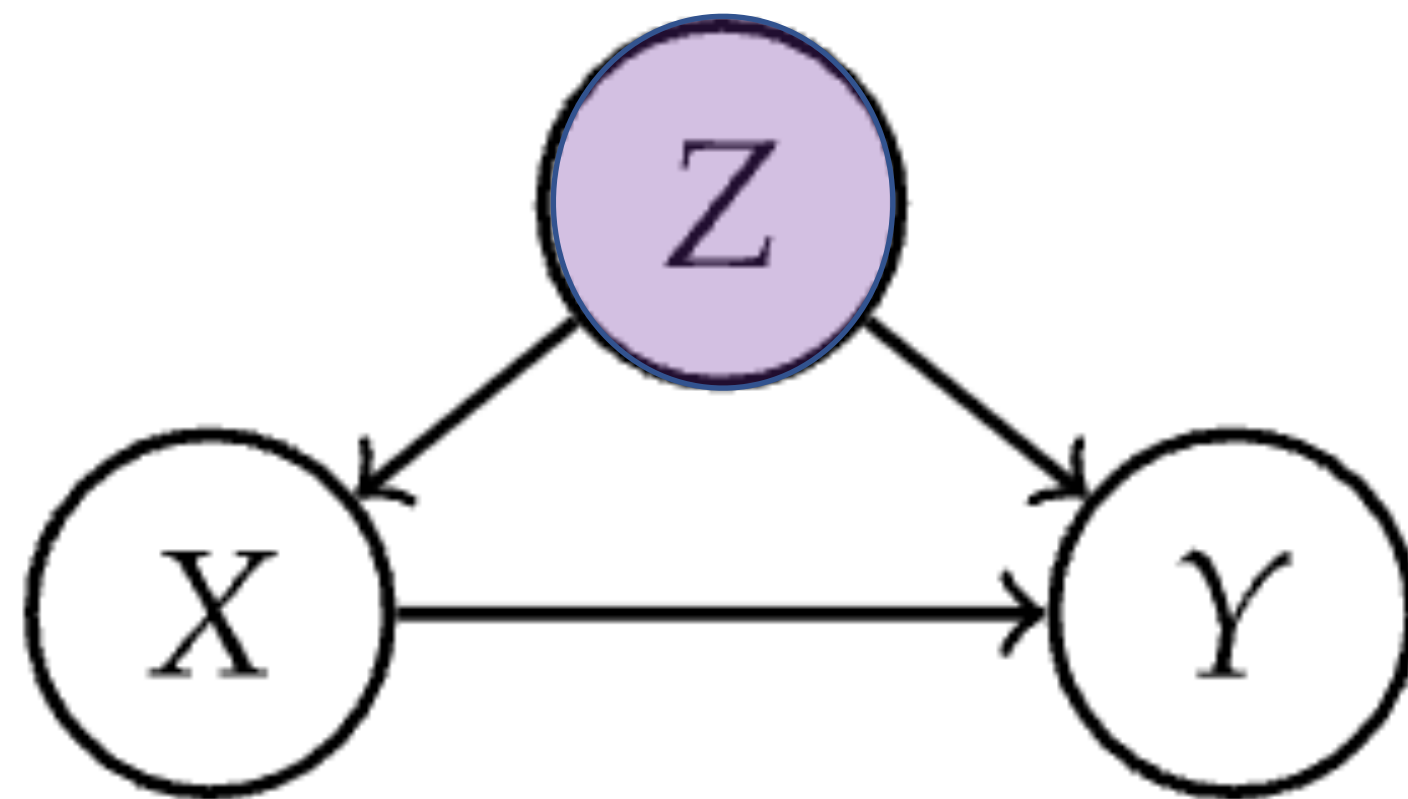
- Кейс 2: Insurance – страховой агент

Вы работаете в страховой компании и хотите предложить клиентам новый продукт – страховку для людей, которые собираются проходить лечение лекарством А. Вам нужно оценить вероятность наступления страхового случая, чтобы потом рассчитать цену продукта.

- Кейс 3: Ученый

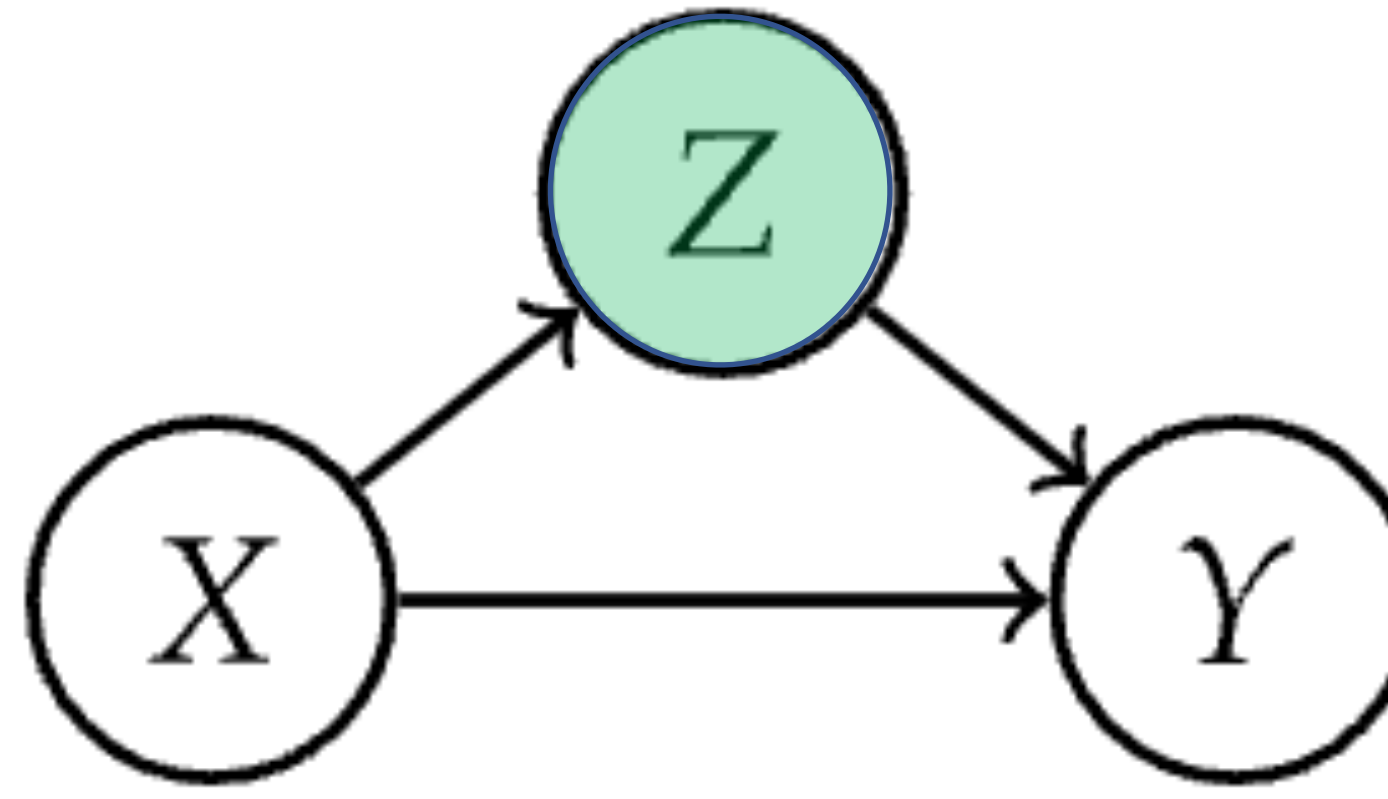
Вы – исследователь и хотите изучить болезнь (почечную недостаточность).

# Типы связей в каузальных графах



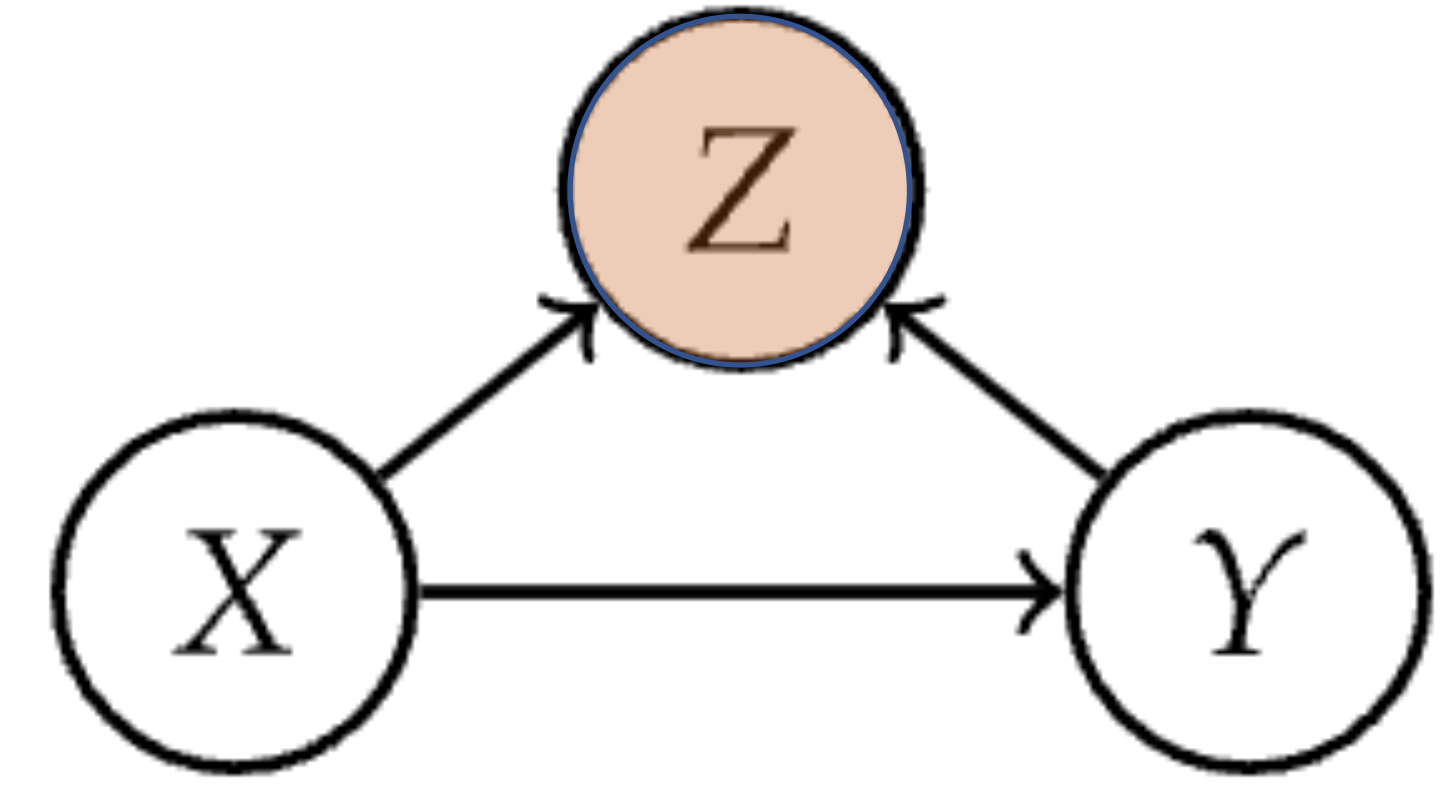
Общая причина

Делает X и Y  
связанными косвенно



Один из путей от X к Y

Не создает confounded  
связи, только causal



Общее следствие

Условие на Z изменяет  
совместное распределение

# Спутанные переменные

- $X$  и  $Y$  называются спутанными, если:

$$P\{Y = y \mid do(X := x)\} \neq P\{Y = y \mid X = x\}$$

- Adjustment формула:

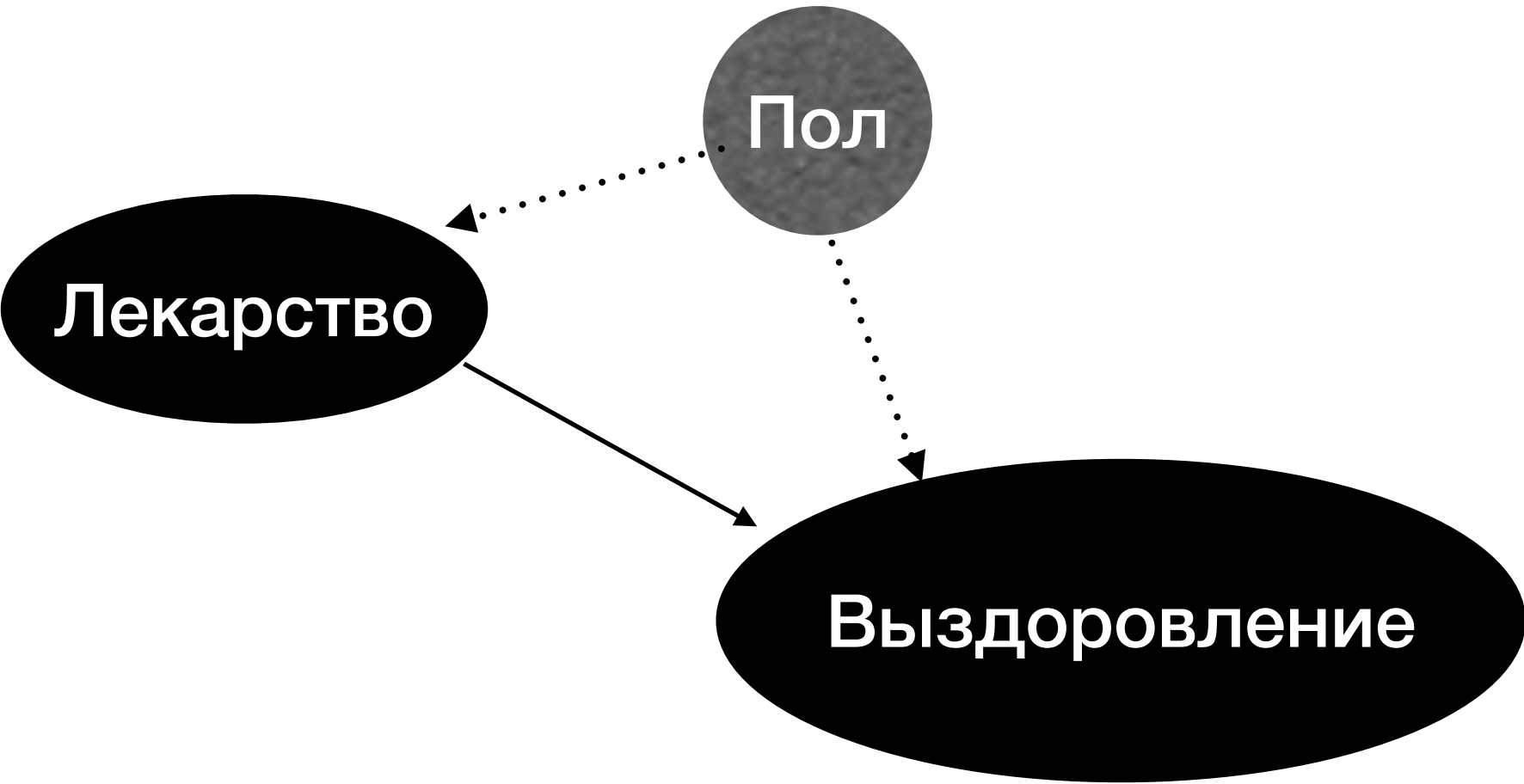
$$P\{Y = y \mid do(X := x)\} = \sum_z P\{Y = y \mid X = x, PA = z\} \cdot P\{PA = z\}$$

$$P\{Y = y \mid X = x\} = \sum_z P\{Y = y \mid X = x, PA = z\} \cdot P\{PA = z \mid X = x\}$$

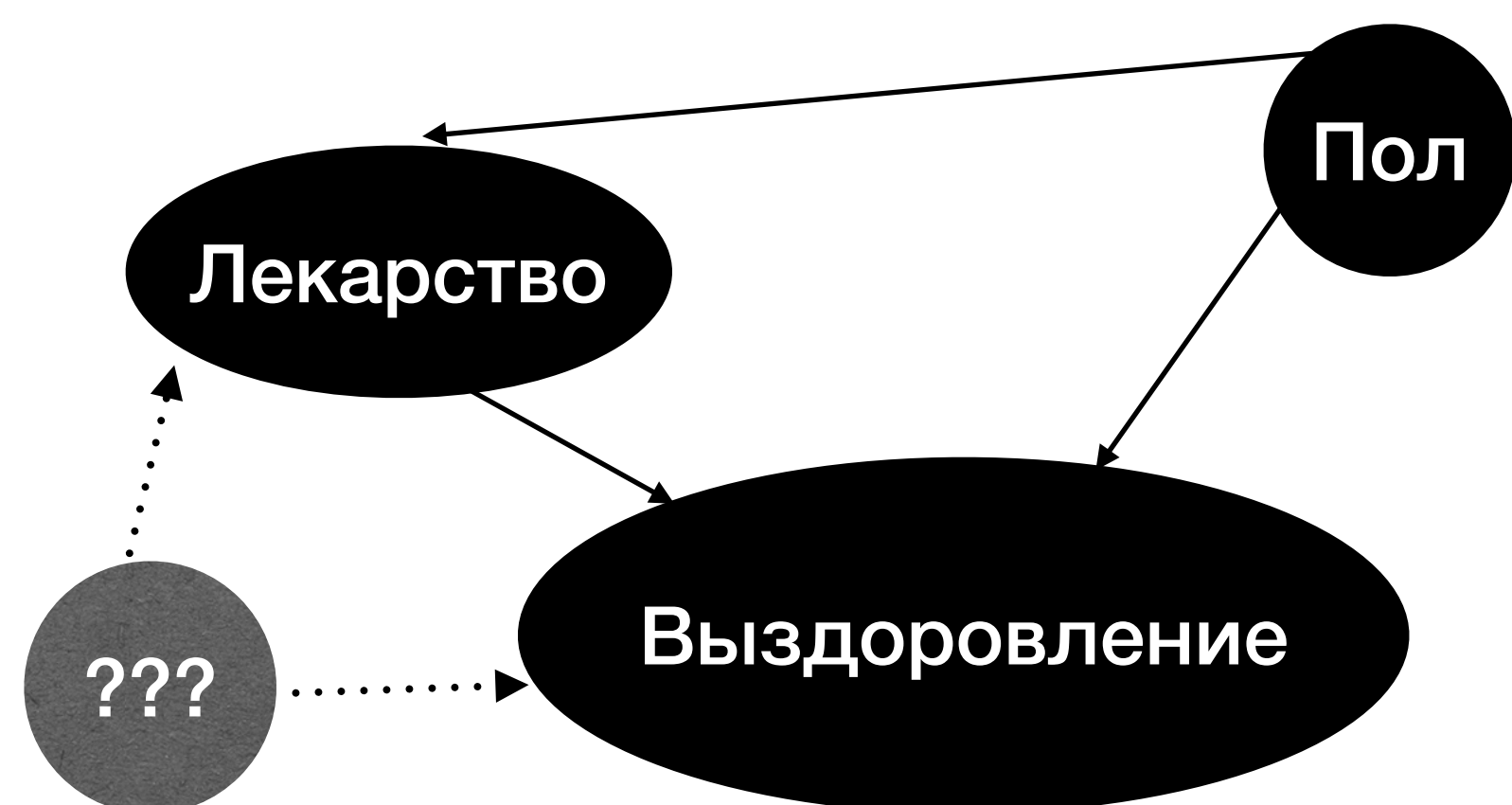


# Пример применения adjustment формулы

Recovery	Drug	No Drug
Total	273/350 (78%)	289/350 (83%)



# Пример применения adjustment формулы



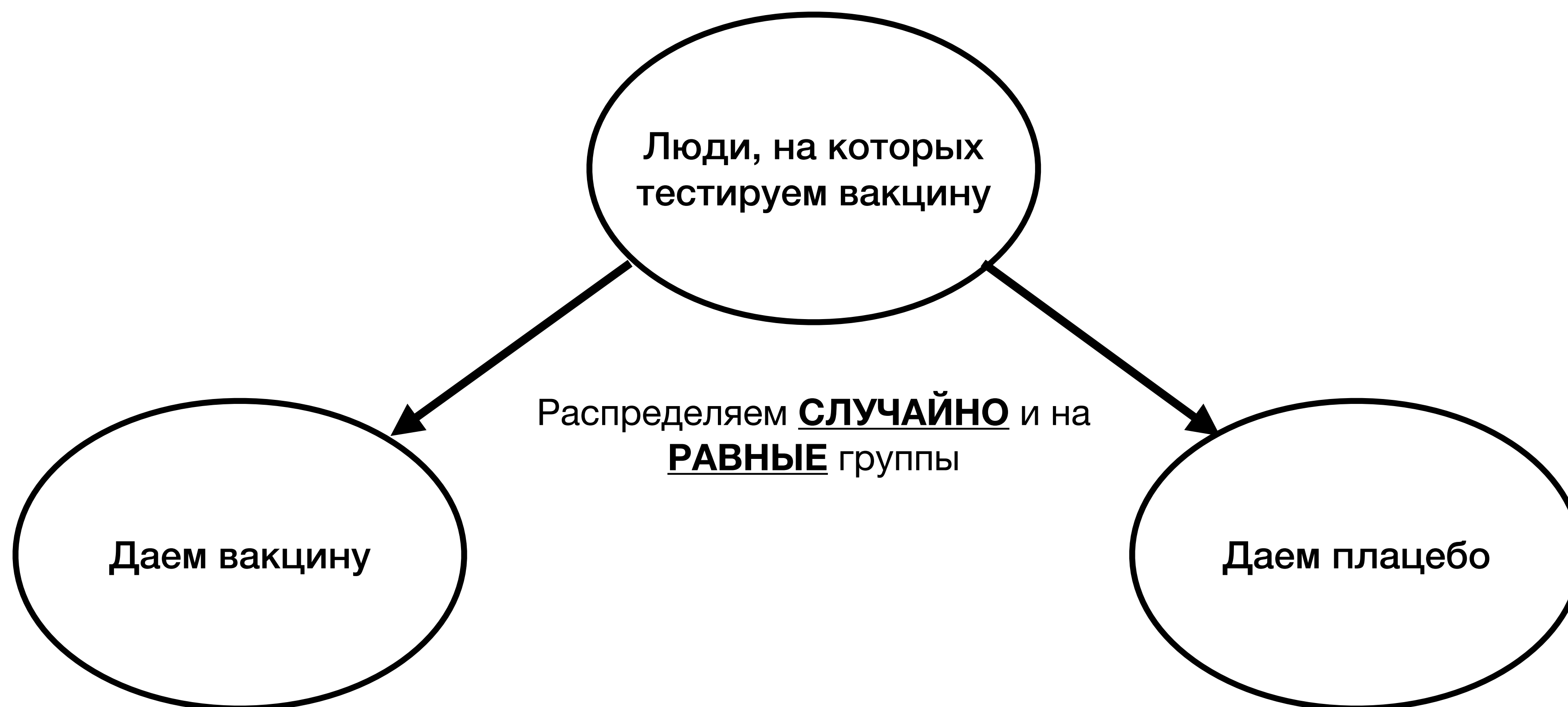
Recovery	Drug	No Drug
Men	81/87 (93%)	234/270 (87%)
Women	192/263 (73%)	55/80 (69%)
Total	273/350 (78%)	289/350 (83%)

$$P(\text{Recovery}|\text{do}(\text{Drug})) = 0.93 \cdot \frac{87 + 270}{700} + 0.73 \cdot \frac{263 + 80}{700} = 0.832$$

$$P(\text{Recovery}|\text{do}(\text{No Drug})) = 0.87 \cdot \frac{87 + 270}{700} + 0.69 \cdot \frac{263 + 80}{700} = 0.7818$$

# Как бороться со спутанными переменными?

## Случайное контролируемое исследование



# Вакцины

- $Y = \begin{cases} Y_T & \text{если дали вакицину} \\ Y_C & \text{если не дали вакцину} \end{cases}$
- Хотим знать  $Y_T, Y_C$  для каждого человека
- Но не получится!



# Разные метрики качества лекарства

- Average Treatment Effect =  $\overline{Y}_T - \overline{Y}_C$

Далее  $Y \in \{0,1\}$

- Odds Ratio =  $\frac{\overline{Y}_T}{1 - \overline{Y}_T} \cdot \frac{1 - \overline{Y}_C}{\overline{Y}_C}$
- Risk Ratio =  $\frac{\overline{Y}_T}{\overline{Y}_C}$
- Effectiveness = 1 - Risk Ratio

# Другой способы борьбы

## Допустимые переменные

- $X$  - допустимая, если

$$P\{Y = y \mid do(T := t)\} = \sum_x P\{Y = y \mid T = t, X = x\} \cdot P\{X = x\}$$

- Сразу получаем способ подсчета do-оператора(если найдем такую переменную)(так можно делать только если  $P\{T = t, X = x\} > 0 \forall x, t$ ):
  1. Собрать данные  $(t_i, y_i, x_i)_{i=1}^n$
  2. Вычислить оценки условных и обычных вероятностей из правой части
  3. Собрать взвешенную сумму

# Другие способы борьбы

## Проблемы предыдущего подхода

- Большая область значения  $X$ 
  - $X$  может отвечать сразу за много параметров: возраст, пол, вес и т.д.
- При увеличении количества признаков область определения увеличивается очень быстро:
  - Увеличивается вычислительная сложность
  - Чтобы добиться хороших оценок нужна большая выборка

Дальше мы будем хотеть найти следующую величину:

$$ATE = \mathbb{E}[Y | do(T := 1)] - \mathbb{E}[Y | do(T := 0)] - \text{average treatment effect}$$



# Сведение к ML

## Propensity scores

- Пусть  $T \in \{0,1\}$ . Обозначим за  $e(x) := \mathbb{E}[T | X = x]$  - propensity score
- Пусть  $X$  - допустимая переменная и  $e(x) \neq 0 \forall x$ . Тогда

$$\mathbb{E}[Y | do(T := 1)] = \mathbb{E} \left[ \frac{YT}{e(X)} \right] \text{ и } \mathbb{E}[Y | do(T := 0)] = \mathbb{E} \left[ \frac{Y(1 - T)}{1 - e(X)} \right]$$

Значит,  $ATE = \mathbb{E} \left[ Y \left( \frac{T}{e(X)} - \frac{1 - T}{1 - e(X)} \right) \right]$  — inverse propensity score weighting.

# Сведение к ML

## Propensity scores

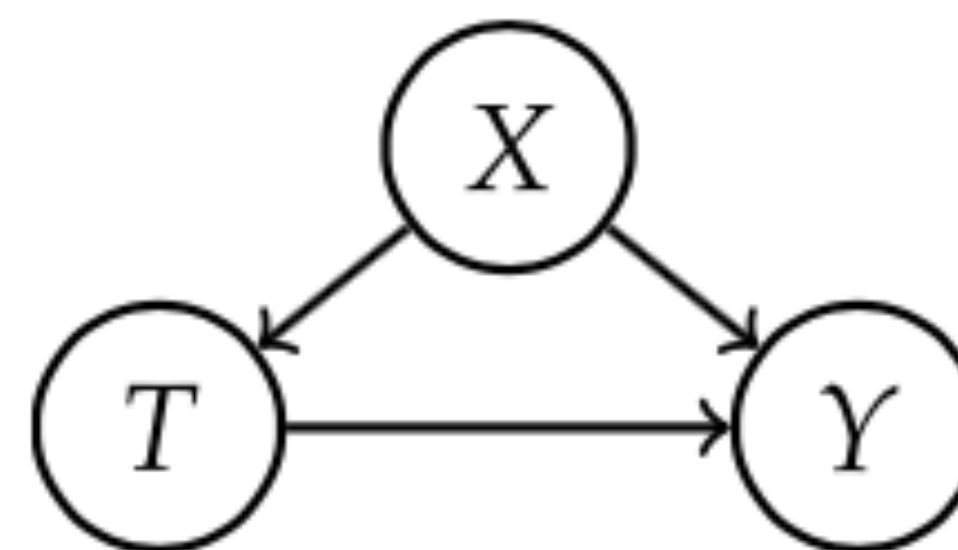
- Делаем пары  $(x_i, e_i) \Rightarrow$  запускаем какую-нибудь модель, которая будет предсказывать  $\hat{e}(x)$ . Будем надеяться, что это оценка получилась хорошей для всех  $x$ .
- Считаем ответ по формуле:

$$\hat{ATE} = \frac{1}{n} \sum_{i=1}^n \frac{t_i y_i}{\hat{e}(x_i)} - \frac{(1 - t_i) y_i}{1 - \hat{e}(x_i)}$$

# Сведение к ML

## Double machine learning

- $Y = \tau T + g(X) + U$
- $T = f(X) + V$
- Мы хотим понять, чему равно  $\tau = ATE$
- $\mathbb{E}[Y|X] = \tau \mathbb{E}[T|X] + g(X)$
- $\tilde{Y} := Y - \mathbb{E}[Y|X] = \tau(T - \mathbb{E}[T|X]) + U =: \tau \tilde{T} + U,$



# Сведение к ML

## Double machine learning

- Находим  $\mathbb{E}[Y|X]$  и  $\mathbb{E}[T|X]$  из  $(X, T, Y)$
- Находим с помощью регрессии  $\hat{\tau} : \hat{Y} = \hat{\tau}\hat{T} + U$

# Еще сведения

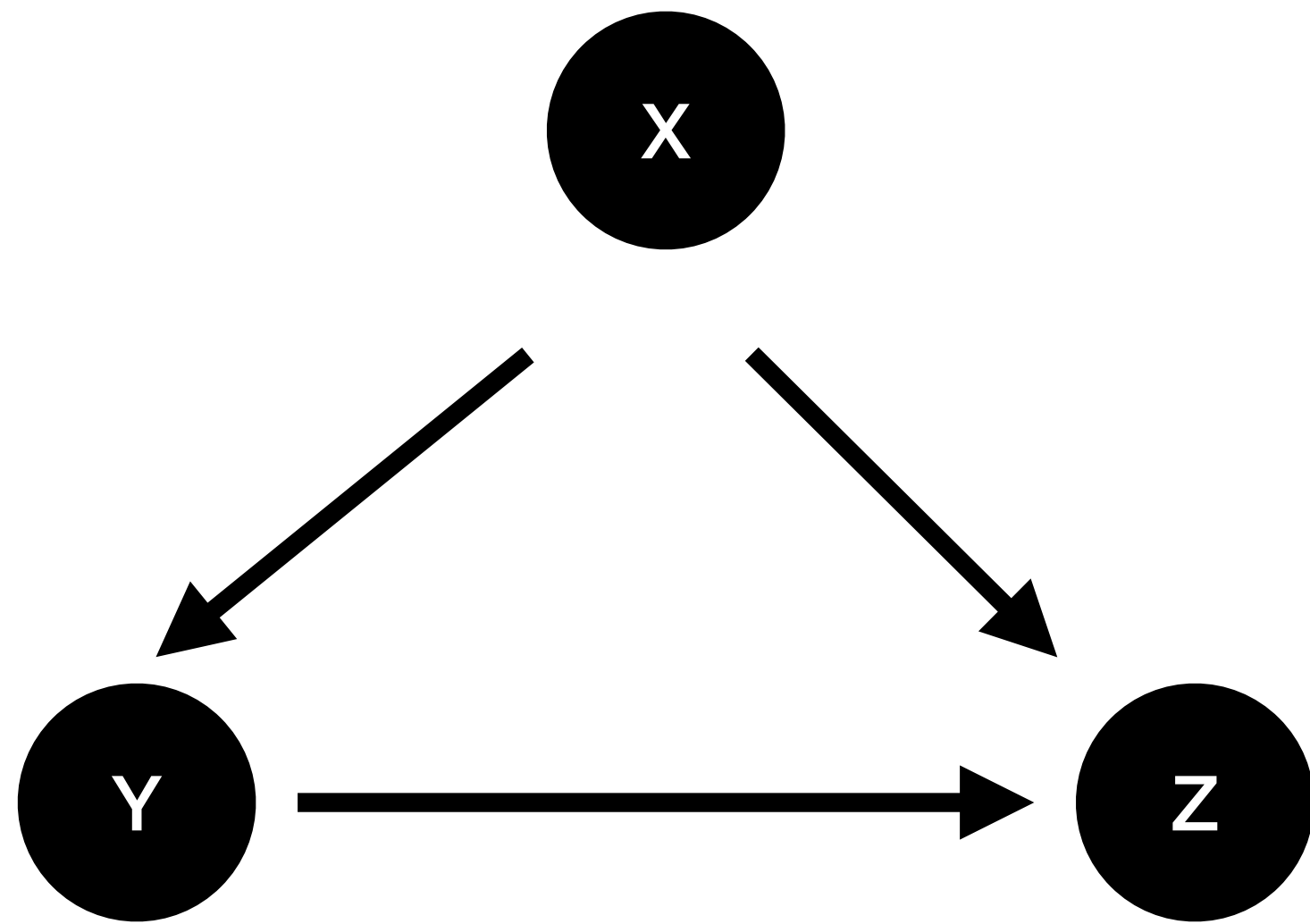
## Heterogeneous treatment effects

- $\tau(x) = \mathbb{E}[Y | do(T := 1), X = x] - \mathbb{E}[Y | do(T := 0), X = x]$  — conditional ATE
- Можно использовать оба предыдущих метода:
  - $\tau(x) = \mathbb{E} \left[ Y \left( \frac{T}{e(X)} - \frac{1 - T}{1 - e(X)} \right) | X = x \right]$
  - $\hat{Y} = \hat{\tau}(X)\hat{T} + U$ , где  $\hat{\tau}(X)$  лежит в некотором классе функций

Model	Predict in i.i.d. setting	Predict under distr. shift/intervention	Answer counter- factual questions	Obtain physical insight	Learn from data
Mechanistic/physical	yes	yes	yes	yes	?
Structural causal	yes	yes	yes	?	?
Causal graphical	yes	yes	no	?	?
Statistical	yes	no	no	no	yes



# Graph Causal Models



+

**Data**

# The 4 key steps

1. **Modeling:** Create a causal graph
2. **Identification:** Formulate what to estimate
3. **Estimation:** Compute the estimate
4. ???

# The 4 key steps of causal inference

1. **Modeling:** Create a causal graph
2. **Identification:** Formulate what to estimate
3. **Estimation:** Compute the estimate
4. **Refutation:** Validate the assumptions

# The 4 key steps of causal inference

1. **Modeling:** Create a causal graph
2. **Identification:** Formulate what to estimate
3. **Estimation:** Compute the estimate
4. **Refutation:** Validate the assumptions

Что самое сложное?

# The 4 key steps of causal inference

1. **Modeling:** Create a causal graph
2. **Identification:** Formulate what to estimate
3. **Estimation:** Compute the estimate
4. **Refutation:** Validate the assumptions

Что самое сложное?

# Intuition about causal graphs

- Assumptions are encoded by *missing edges*, and *direction of edges*

# Intuition about causal graphs

- Assumptions are encoded by *missing edges*, and *direction of edges*
- Graph cannot be learn from data alone



# Intuition about causal graphs

- Assumptions are encoded by *missing edges*, and *direction of edges*
- Graph cannot be learn from data alone
- No test, no cross-validation

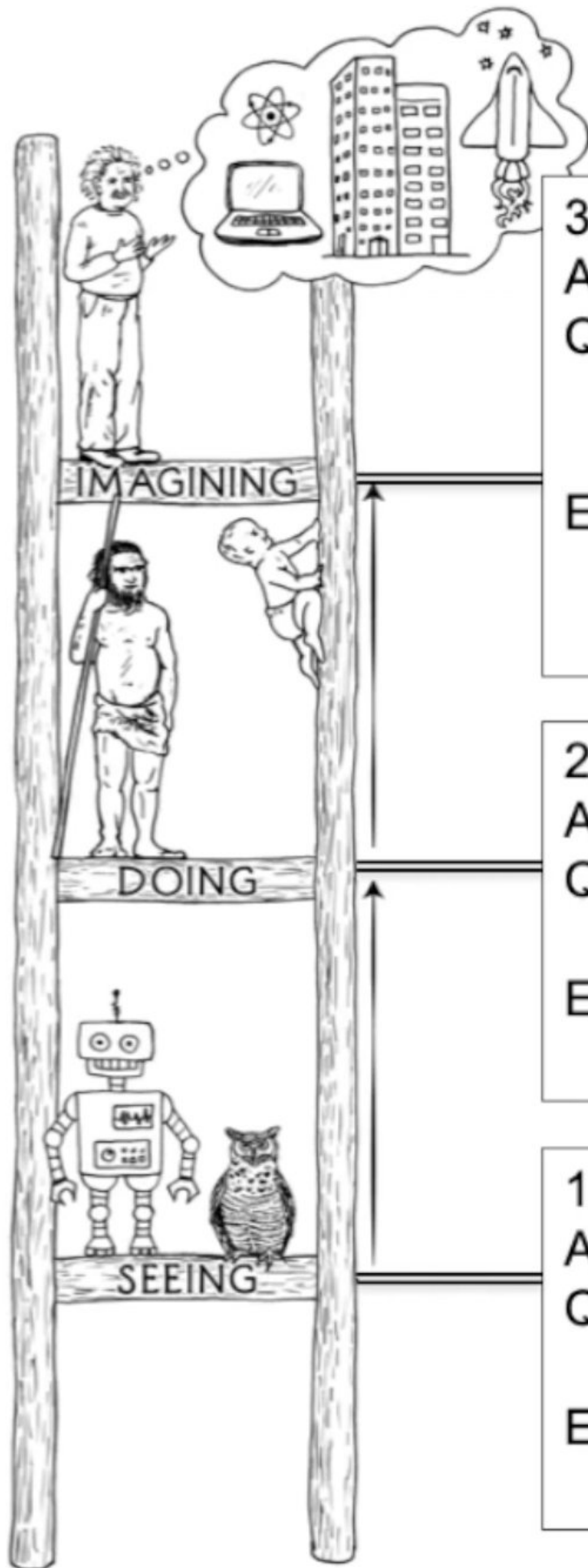
# Intuition about causal graphs

- Assumptions are encoded by *missing edges*, and *direction of edges*
- Graph cannot be learn from data alone
- No test, no cross-validation
- Do for as many assumptions as possible

# Counterfactuals

- **Interventional question:**  
How does the probability of heart failure change if we convince a patient to exercise regularly?
- **Counterfactual question:**  
Would a given patient have suffered heart failure if they had started exercising a year earlier?

# 3-LEVEL HIERARCHY



## 3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done . . . ? Why?*

(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache?  
Would Kennedy be alive if Oswald had not killed him? What if I had not smoked the last 2 years?

## 2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do . . . ? How?*

(What would Y be if I do X?)

EXAMPLES: If I take aspirin, will my headache be cured?  
What if we ban cigarettes?

## 1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see . . . ?*

(How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?  
What does a survey tell us about the election results?

# Example

Randomly choose  $R$  (route).

$Y$  - is bad traffic

If  $B$  (bad day) == 1 then  $Y = 1$ .

Else with  $U_R \sim B(1/2)$  there is an accident.

$$\begin{cases} R \sim B(1/2) \\ W \sim B(1/2) \\ U_0, U_1 \sim B(1/2) \end{cases}$$

# Example

Randomly choose  $R$  (route).

If  $B$  (bad traffic day)  $== 1$  then  $Y = 1$ .

Else with  $U_R \sim B(1/2)$  there is an accident.

$$\begin{cases} R \sim B(1/2) \\ B \sim B(1/2) \\ B_0, B_1 \sim B(1/2) \end{cases}$$

Suppose  $R = 1$ ,  $Y = 1$ .

Would we have been better off taking the alternative route this morning?

# Example

$$Y = R \max(B, B_1) + (1 - R) \max(B, B_0)$$

How to find?

Do-operator  $P(Y \mid do\{R = 0\})$  and we get  $(1 / 2) * (1 / 2) = 1 / 4$ .



# Example

B	B_1
0	1
1	0
1	1

$$Y = R \max(B, B_1) + (1 - R) \max(B, B_0)$$

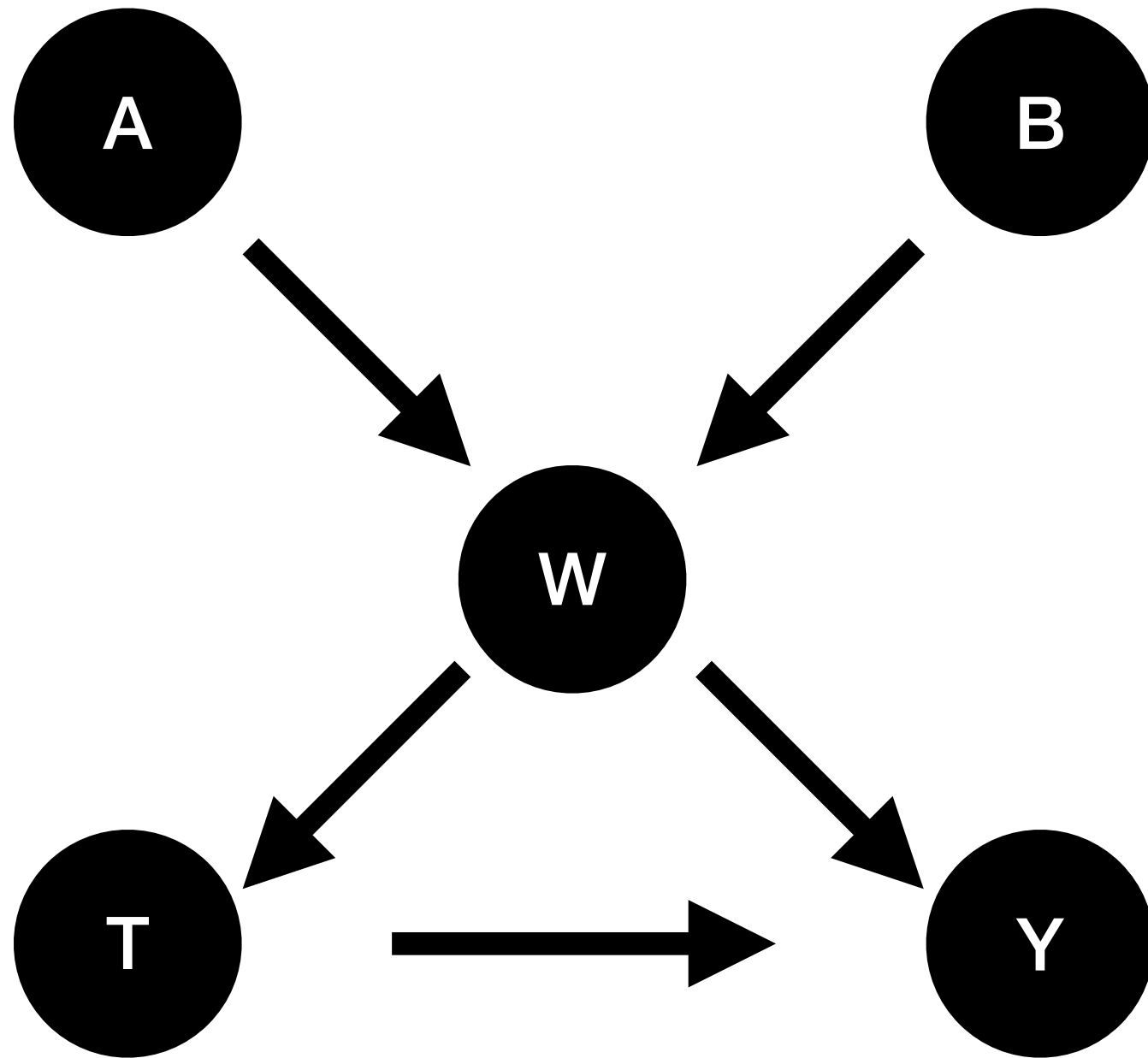
We get new distribution  $B'$ . Then we apply Do-operator:

$$Y = \max(B', B_0) \implies P(Y = 0 \mid do(R = 0)) = 1/3 \cdot 1/2 = 1/6$$

# Connection with ML

- Recommendation systems
- Robustness and Generalisation
- Reinforcement Learning

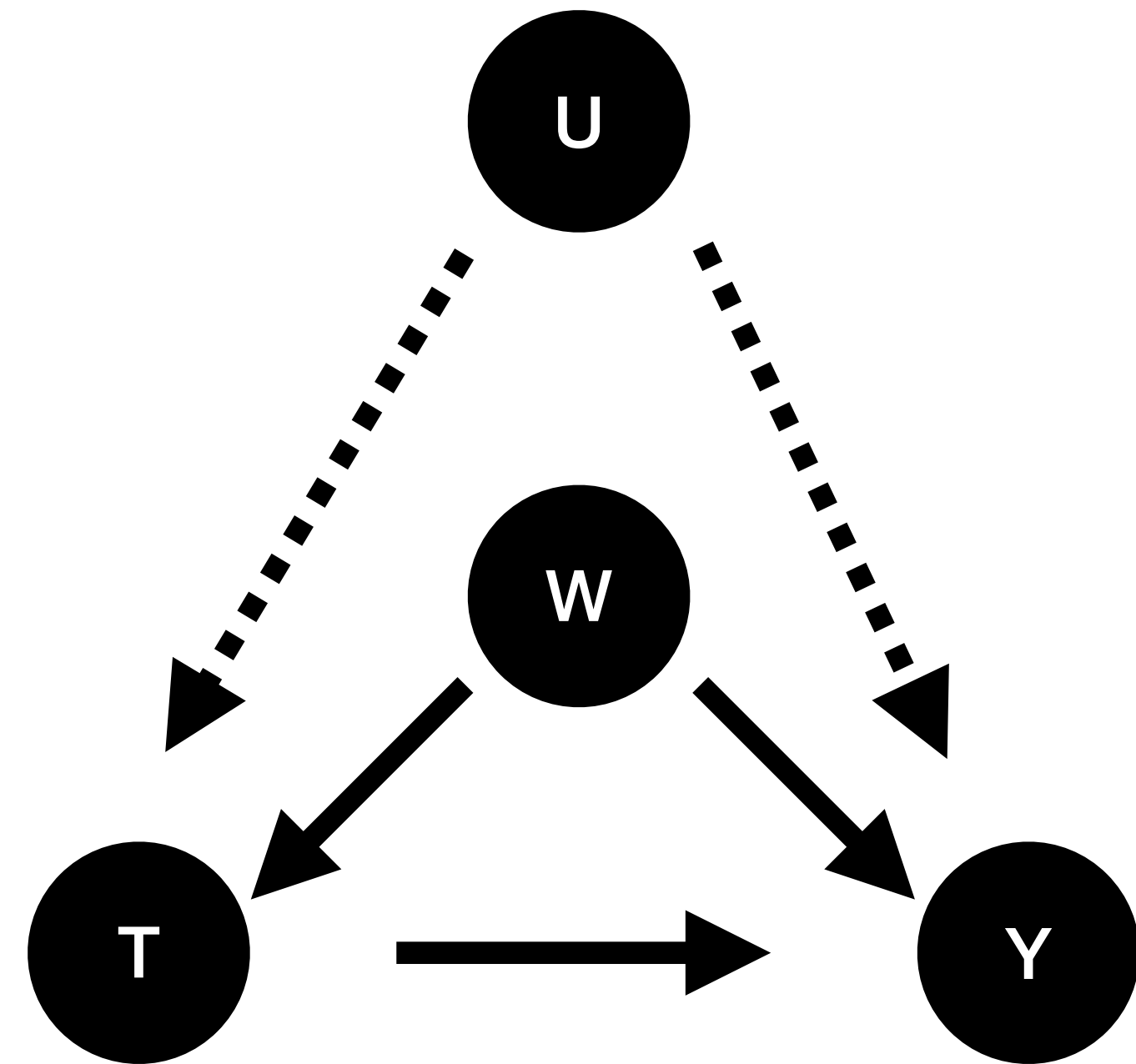
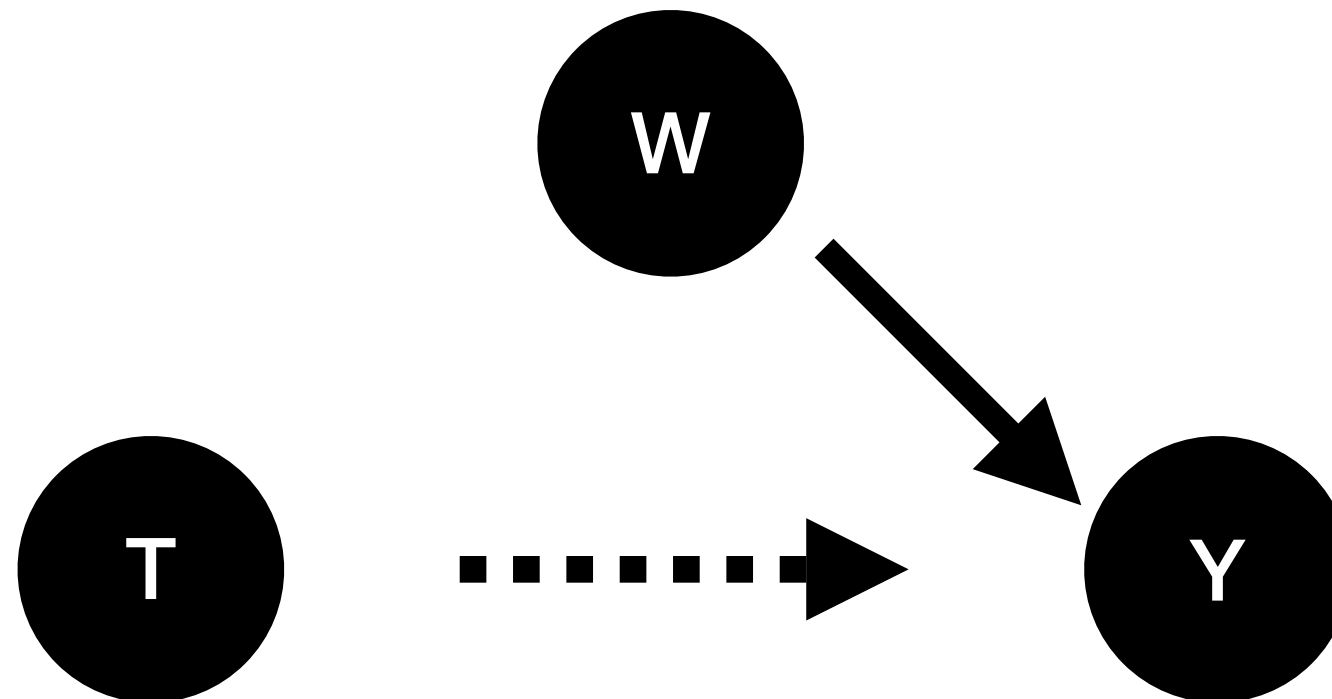
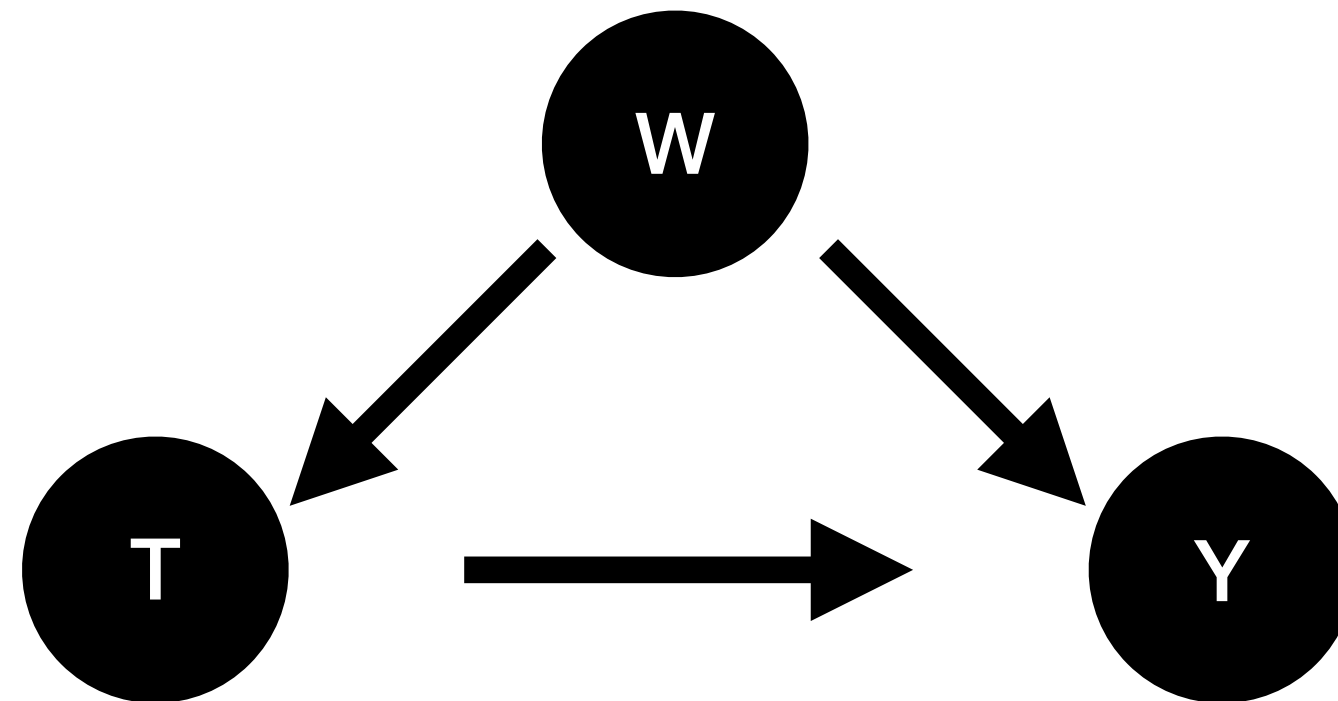
# Some tests



$$A \perp\!\!\!\perp B$$

$$A \perp\!\!\!\perp T \mid W$$

$$B \perp\!\!\!\perp T \mid W$$



# СПИСОК ИСТОЧНИКОВ

- <https://mlstory.org/causal.html>
- <https://mlstory.org/causal-practice.html>
- <https://arxiv.org/abs/2102.11107>
- <https://mitpress.mit.edu/books/elements-causal-inference> (ch. 1-7)
- <https://youtu.be/LALfQStONEc>
- <https://causalinference.gitlab.io/book/> (chapter 1)
- [https://www.youtube.com/watch?v=HOgx\\_SBBzn0](https://www.youtube.com/watch?v=HOgx_SBBzn0)
- <https://www.youtube.com/watch?v=RtxTpOb8e4>