Risk ratio, odds ratio, risk difference…

Which causal measure is easier to generalize?

Bénédicte Colnet, Ph.D. student at Inria (Soda & PreMeDICaL teams)

EuroCIM, Oslo, April 20th

Julie Josse
Missing values & causal inference

Gaël Varoquaux
ML & co-founder of scikit-learn

Erwan Scornet
Random forest & missing values

ArXiv

Inria

Inserm

A variety of causal measures

Clinical example from Cook and Sackett (1995)

Randomized Controlled Trial (RCT),

- **Y** the observed binary outcome (stroke after 5 years)
- **A** binary treatment assignment
- **X** baseline covariates

RCT's findings

11.1% stroke in control, versus 6.7% in treated

Usually referring to an *effect*, is related to how one *contrasts* those two

e.g. Ratio = 6.7/11.1 = 0.6 **or** Diff = - 0.04

A variety of causal measures

Note that for binary Y,
E[Y(a)] = P(Y=1 | A=a)

Clinical example from Cook and Sackett (1995)

Randomized Controlled Trial (RCT),

- **Y** the observed binary outcome (stroke after 5 years)
- **A** binary treatment assignment
- **X** baseline covariates

Potential outcomes framework → $\mathbb{E}\left[Y^{(0)}\right]$, $\mathbb{E}\left[Y^{(1)}\right]$

RCT's findings

11.1% stroke in control, versus 6.7% in treated

Usually referring to an *effect*, is related to how one *contrasts* those two

e.g. Ratio = 6.7/11.1 = 0.6 **or** Diff = - 0.04

Count the stroke

$$\tau_{RR} = \frac{\mathbb{E}\left[Y^{(1)}\right]}{\mathbb{E}\left[Y^{(0)}\right]}$$

Count the non-stroke

$$\tau_{SR} = \frac{1 - \mathbb{E}\left[Y^{(1)}\right]}{1 - \mathbb{E}\left[Y^{(0)}\right]}$$

$$\tau_{RD} = \mathbb{E}\left[Y^{(1)}\right] - \mathbb{E}\left[Y^{(0)}\right]$$

Risk Difference

$$\tau_{NNT} = \tau_{RD}^{-1}$$

Number Needed to Treat

Odds Ratio

$$\tau_{OR} = \frac{\mathbb{E}[Y^{(1)}]}{1 - \mathbb{E}[Y^{(1)}]} \left(\frac{1 - \mathbb{E}[Y^{(0)}]}{1 - \mathbb{E}[Y^{(0)}]} \right)^{-1}$$

A variety of causal measures

Continuing the clinical example

X = 1 <–> high baseline risk

	τ_{RD}	τ_{RR}	τ_{SR}	τ_{NNT}	τ_{OR}
All (P_s)	-0.0452	**0.6**	1.05	22	0.57
X = 1	-0.006	**0.6**	1.01	167	0.6
X = 0	-0.08	**0.6**	1.1	13	0.545

Computed from Cook & Sackett (1995)

Marginal effects τ

Conditional effects $\tau(x)$

``Treated group has 0.6 times the risk of having a stroke outcome when compared with the placebo.''* ***or*** *``The Number Needed to Treat is 22.'' ***or*** *``Effect is stronger on subgroup X=0 but not on the ratio scale.''*

— leading to different impressions and heterogeneity patterns

The age-old question of how to report effects

Source: Wikipedia

`` We wish to decide whether we shall count the failures or the successes and whether we shall make relative or absolute comparisons"

— Mindel C. Sheps, New England Journal of Medicine, in 1958

The choice of the measure is still actively discussed

e.g. Spiegelman and VanderWeele, 2017; Baker and Jackson, 2018; Feng et al., 2019; Doi et al., 2022; Xiao et al., 2021, 2022; Huitfeldt et al., 2021; Lapointe-Shaw et al., 2022; Liu et al., 2022 …

— CONSORT guidelines recommend to report all of them

A desirable property: collapsibility

i.e. population's effect is equal to a weighted sum of local effects

Discussed in Greenland, 1987; Hernàn et al. 2011; Huitfeldt et al., 2019; Daniel et al., 2020; Didelez and Stensrud, 2022 and many others.

A very famous example: the Simpson paradox

Marginal effect bigger than subgroups' effects

(a) Overall population, $\tau_{\text{OR}} \approx 0.26$

	Y=0	Y=1
A=1	1005	95
A=0	1074	26

(b) $\tau_{\text{OR}|F=1} \approx 0.167$ and $\tau_{\text{OR}|F=0} \approx 0.166$

F= 1	Y=0	Y=1
A=1	40	60
A=0	80	20

F=0	Y=0	Y=1
A=1	965	35
A=0	994	6

Toy example inspired from Greenland (1987).

— Unfortunately, not all measures are collapsible

Collapsibility and formalism

- Different definitions of collapsibility in the literature

- We propose three definitions encompassing previous works

1. Direct collapsibility $\mathbb{E}\left[\tau(X)\right] = \tau$

2. Collapsibility $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\,\tau(X)\right] = \tau,$ **with** $w \geq 0$**, and** $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\right] = 1$

3. Logic-respecting $\tau \in \left[\min_x(\tau(x)), \max_x(\tau(x))\right]$

e.g RR is collapsible, with

$$\mathbb{E}\left[\tau_{RR}(X)\frac{\mathbb{E}\left[Y^{(0)} \mid X\right]}{\mathbb{E}\left[Y^{(0)}\right]}\right] = \tau_{RR}$$

Collapsibility and formalism

- Different definitions of collapsibility in the literature

- We propose three definitions encompassing previous works

1. Direct collapsibility $\mathbb{E}\left[\tau(X)\right] = \tau$

2. Collapsibility $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\,\tau(X)\right] = \tau,$ **with** $w \geq 0$**, and** $\mathbb{E}\left[w(X, P(X, Y^{(0)}))\right] = 1$

3. Logic-respecting $\tau \in \left[\min_x(\tau(x)), \max_x(\tau(x))\right]$

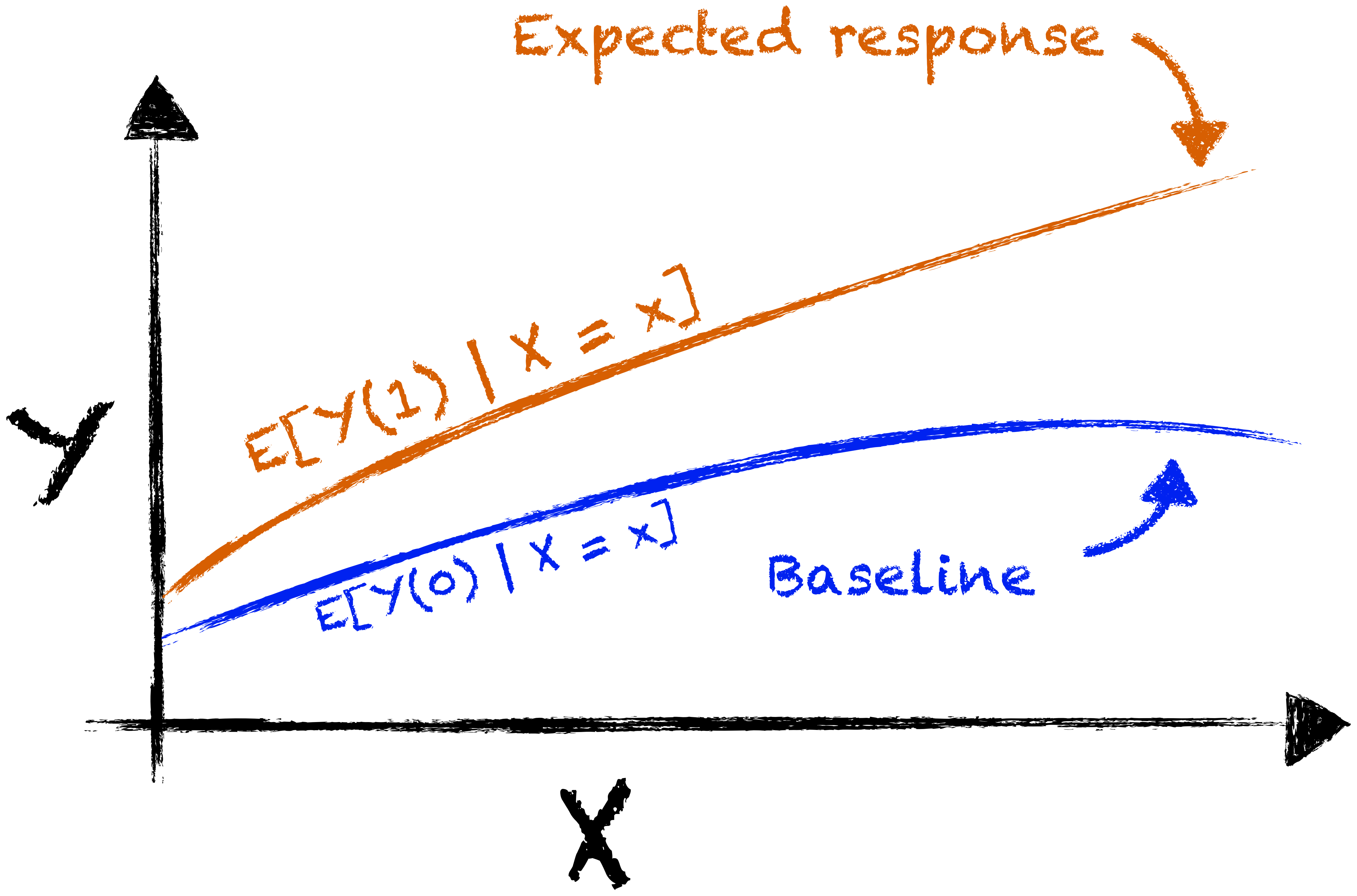
e.g RR is collapsible, with

$$\mathbb{E}\left[\tau_{RR}(X)\frac{\mathbb{E}\left[Y^{(0)} \mid X\right]}{\mathbb{E}\left[Y^{(0)}\right]}\right] = \tau_{RR}$$

Measure	Collapsible	Logic-respecting
Risk Difference (RD)	Yes	Yes
Number Needed to Treat (NNT)	No	Yes
Risk Ratio (RR)	Yes	Yes
Survival Ratio (SR)	Yes	Yes
Odds Ratio (OR)	No	No

Through the lens of non parametric generative models

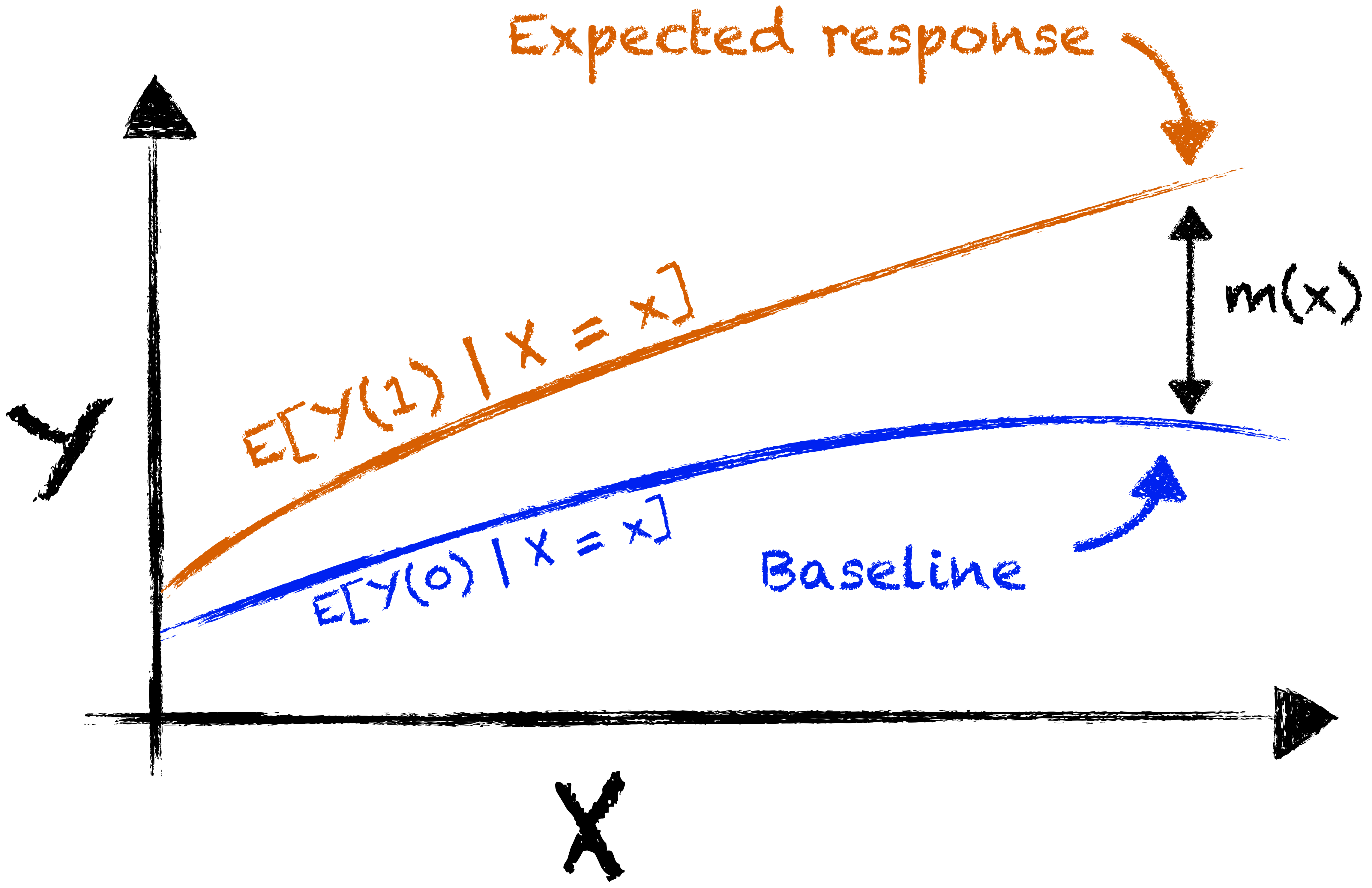
For Y continuous,



(*) This only assumes that conditional expected responses are defined for every x

Through the lens of non parametric generative models

For Y continuous,



(*) This only assumes that conditional expected responses are defined for every x

Lemma*

There exist two functions b(.) and m(.) such that,

$$\mathbb{E}\left[Y^{(a)} \mid X\right] = b(X) + a\, m(X)$$

Additivity

Spirit of Robinson's decomposition (1988), further developed in Nie et al. 2020

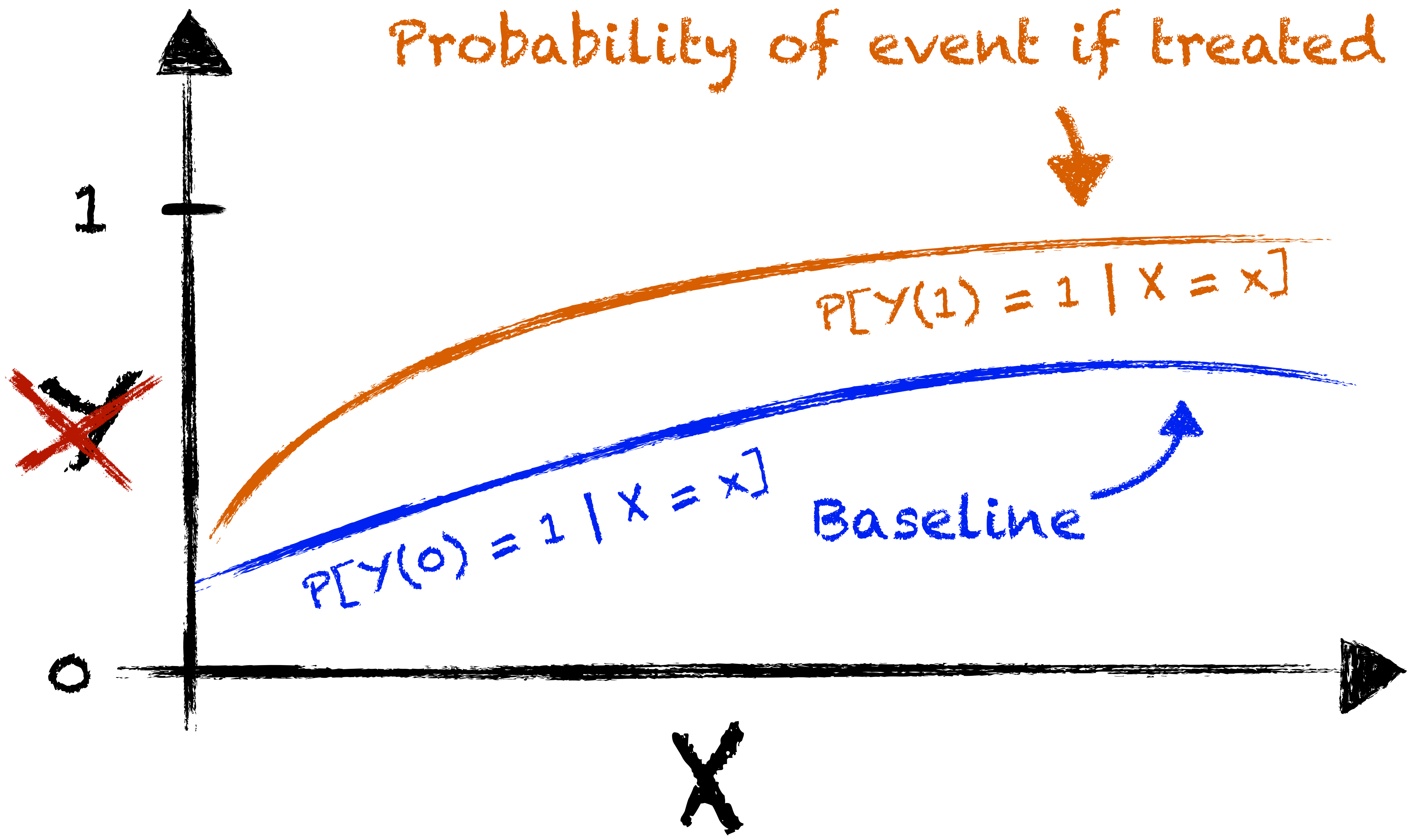
Linking generative functions with measures

$$\tau_{RR}(x) = 1 + m(x)/b(x)$$ Entanglement

$$\tau_{RD}(x) = m(x)$$ No entanglement

Through the lens of non parametric generative models

For Y binary,



Lemma

There exist two functions b(.) and m(.) such that,

$$\mathbb{E}\left[Y^{(a)} \mid X\right] = b(X) + a\, m(X)$$

Additivity

Adapted Lemma

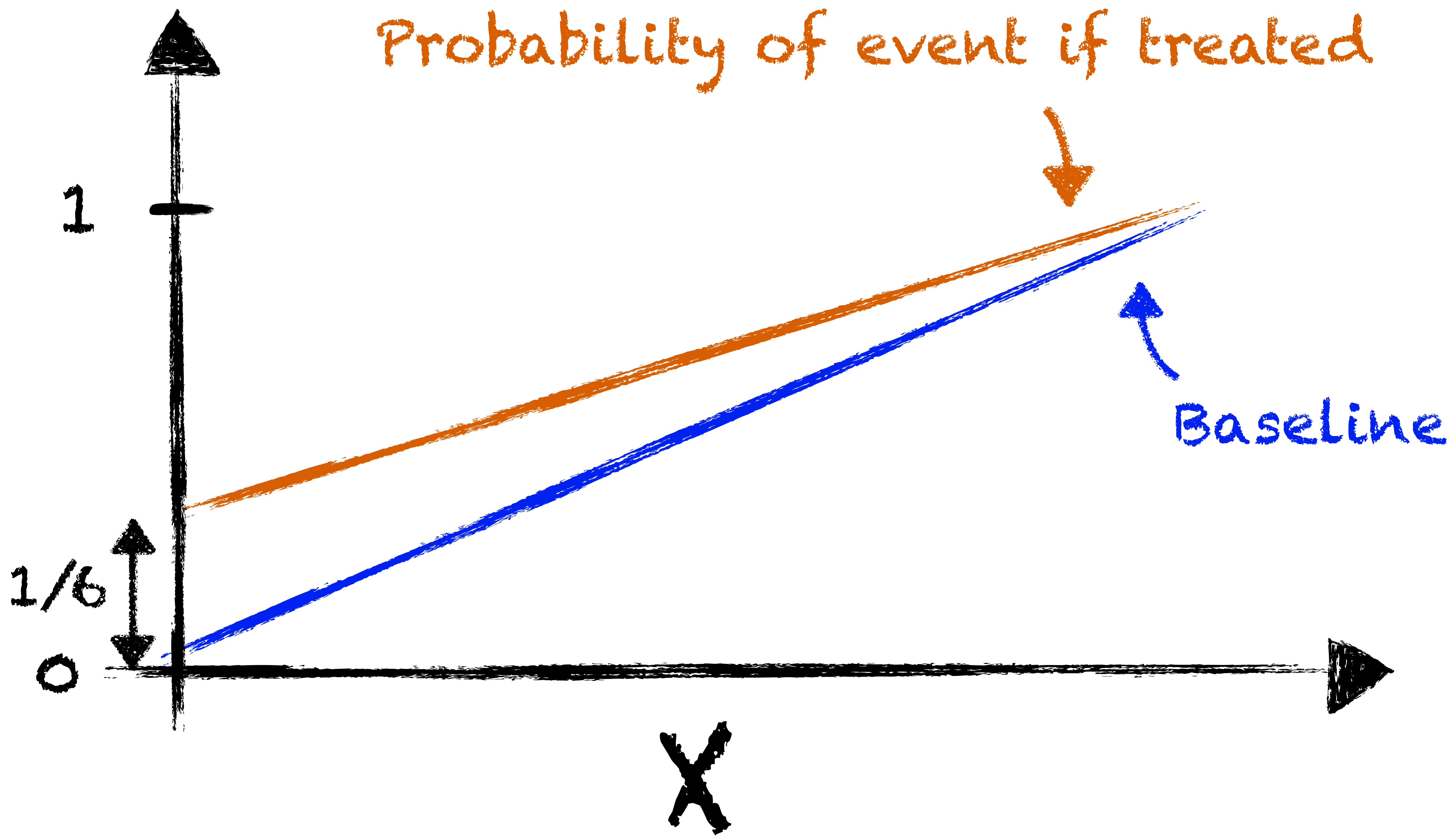
There exist two functions b(.) and m(.) such that,

$$\ln\left(\frac{\mathbb{P}(Y^{(a)} = 1 \mid X)}{\mathbb{P}(Y^{(a)} = 0 \mid X)}\right) = b(X) + a\, m(X)$$

The example of the Russian roulette

Harmful

For Y binary,

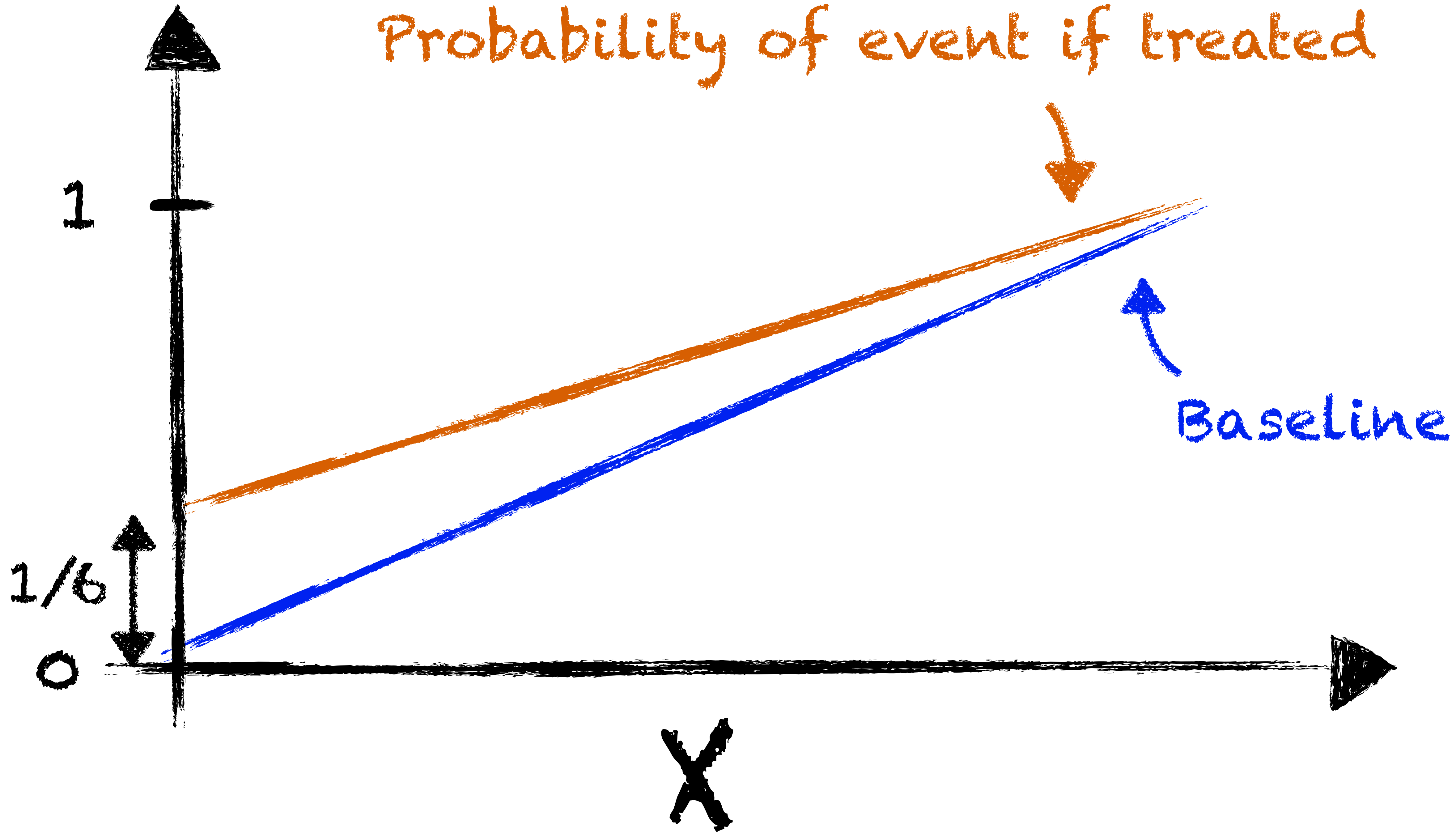


Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

The example of the Russian roulette

Harmful

For Y binary,



Lemma

There exist two functions b(.) and m(.) such that,

$$\mathbb{P}\left[Y^{(a)} = 1 \mid X\right] = b(X) + a\left(1 - b(X)\right)m(X)$$

Simple additivity is not possible anymore

Linking generative functions with measures

$$\tau_{RD}(x) = (1 - b(x))m(x)$$ Entanglement

$$\tau_{SR}(x) = 1 - m(x)$$ No entanglement

Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

Extension to all effect types (harmful and beneficial)

Considering a binary outcome, assume that

$$\forall x \in \mathbb{X}, \forall a \in \{0,1\}, \quad 0 < p_a(x) < 1, \quad \textbf{where } p_a(x) := \mathbb{P}\left[Y^{(a)} = 1 \mid X = x\right]$$

Assumptions

Introducing,

$$m_g(x) := \mathbb{P}\left[Y^{(1)} = 0 \mid Y^{(0)} = 1,\, X = x\right] \quad \textbf{and} \quad m_b(x) := \mathbb{P}\left[Y^{(1)} = 1 \mid Y^{(0)} = 0,\, X = x\right],$$

allows to have,

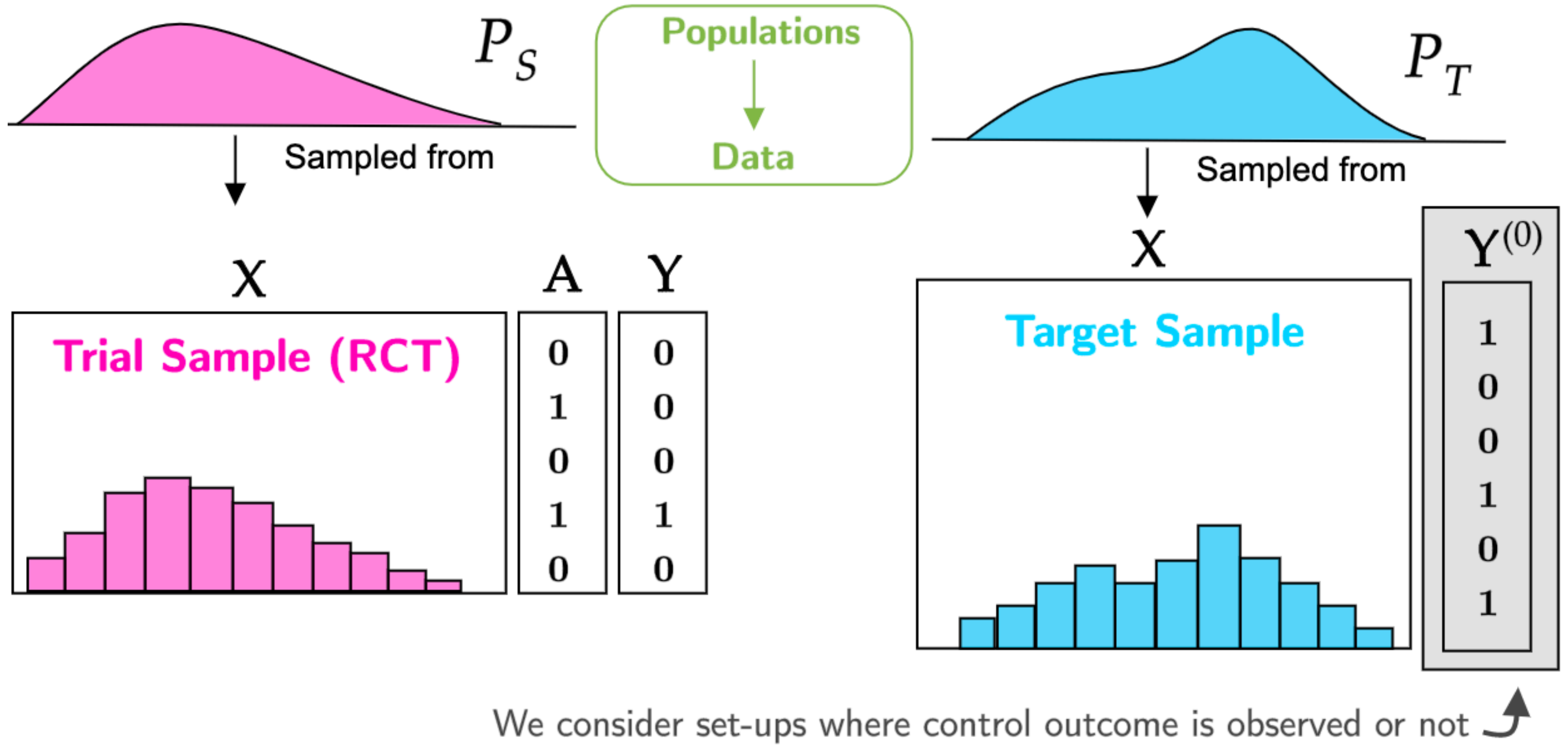
$$\mathbb{P}\left[Y^{(a)} = 1 \mid X = x\right] = b(x) + a\left(\left(1 - b(x)\right) m_b(x) - b(x)\, m_g(x)\right), \quad \textbf{where } b(x) := p_0(x)\,.$$

More events

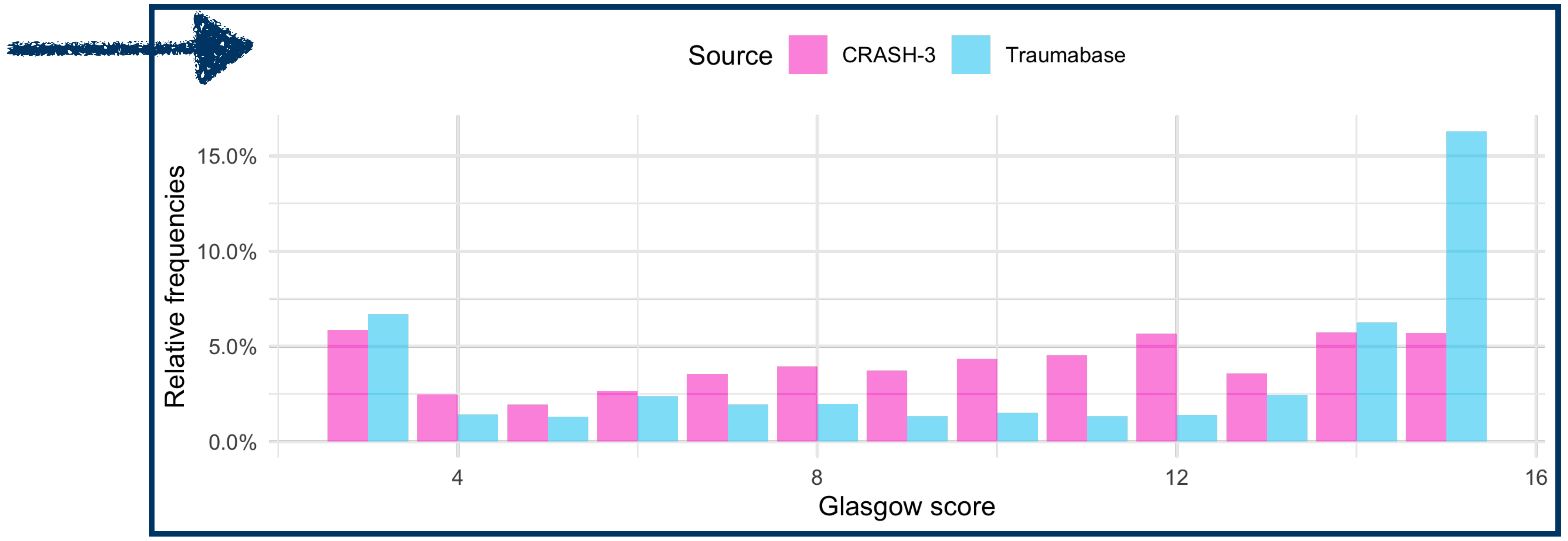
Less events

Generalizability

i.e. transport trial findings to a target population $\hat{\tau}_{RCT} \longrightarrow \hat{\tau}_{Target}$



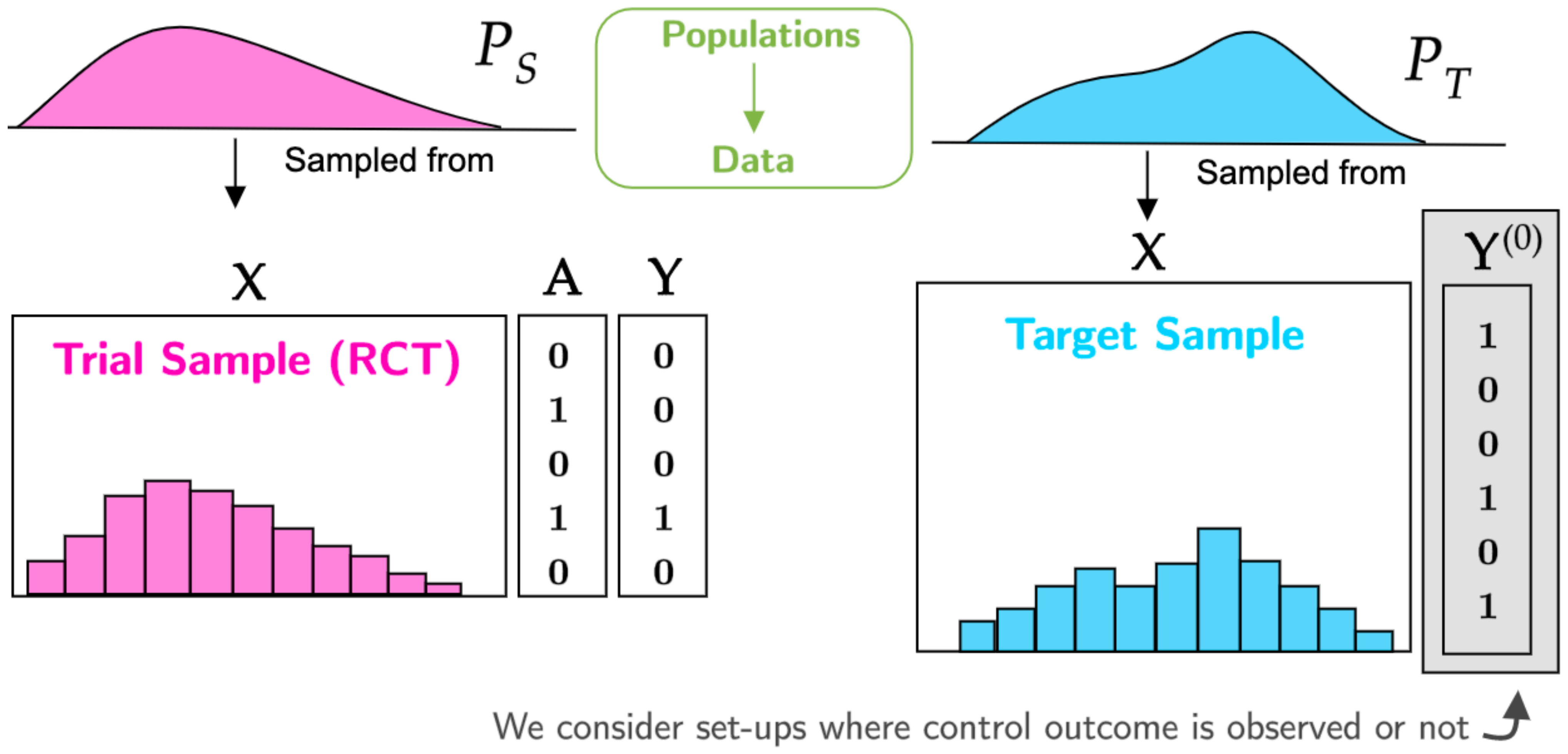
A real-world example



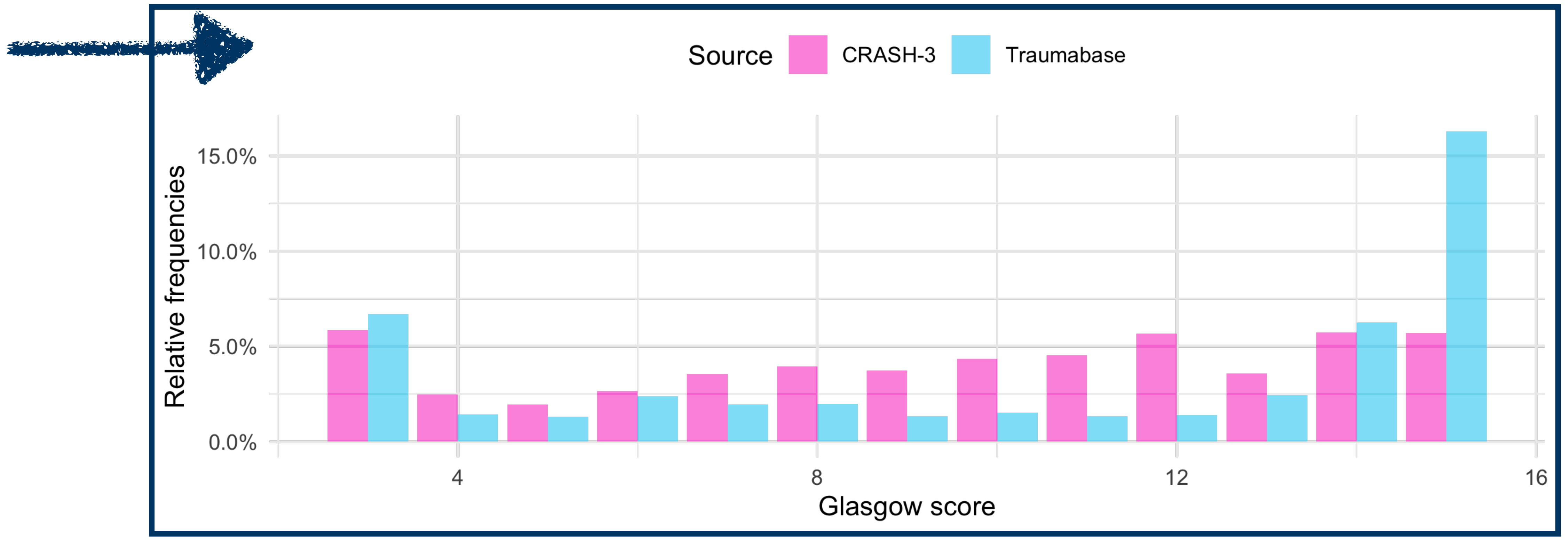
What would be the effect if individuals where sampled in target population?

Generalizability

i.e. transport trial findings to a target population $\hat{\tau}_{RCT} \longrightarrow \hat{\tau}_{Target}$



A real-world example



State-of-the-art

- Ideas present in epidemiological books (Rothman & Greenland, 2000)
- Foundational work from Stuart et al. 2010 and Pearl & Barenboim 2011
- Currently flourishing field with IPW, G-formula, and doubly-robust estimators

Focus on generalizing the difference

Two methods, two assumptions

S is the indicator of population's membership

Generalizing	Conditional potential outcomes	Local effects
Assumptions for RD	$\{Y^{(0)}, Y^{(1)}\} \perp\!\!\!\perp S \mid X$	$Y^{(1)} - Y^{(0)} \perp\!\!\!\perp S \mid X$
Unformal	All shifted prognostic covariates	All shifted treatment effect modifiers Less covariates if homogeneity
Identification	$\mathbb{E}^T\left[Y^{(a)}\right] = \mathbb{E}^T\left[\mathbb{E}^R\left[Y^{(a)} \mid X\right]\right]$	$\tau^T = \mathbb{E}\left[w(X, Y^{(0)})\tau^R(X)\right]$ Possible only if collapsible!

— Depending on the assumptions, either conditional outcome or local treatment effect can be generalised

Generalizing local effect, for a binary Y and a beneficial effect

i.e. reducing number of events

Estimate using trial sample

$$\mathbb{E}\left[\tau_{RR}(X)\,\frac{\mathbb{E}\left[Y^{(0)} \mid X\right]}{\mathbb{E}\left[Y^{(0)}\right]}\right] = \tau_{RR}$$

Estimate using target sample

$$\tau_{RR}(x) = 1 - m_g(x)$$

Thanks to the generative model,
only depends on covariates in m(X)

A toy simulation

Introducing heterogeneities in the Russian roulette

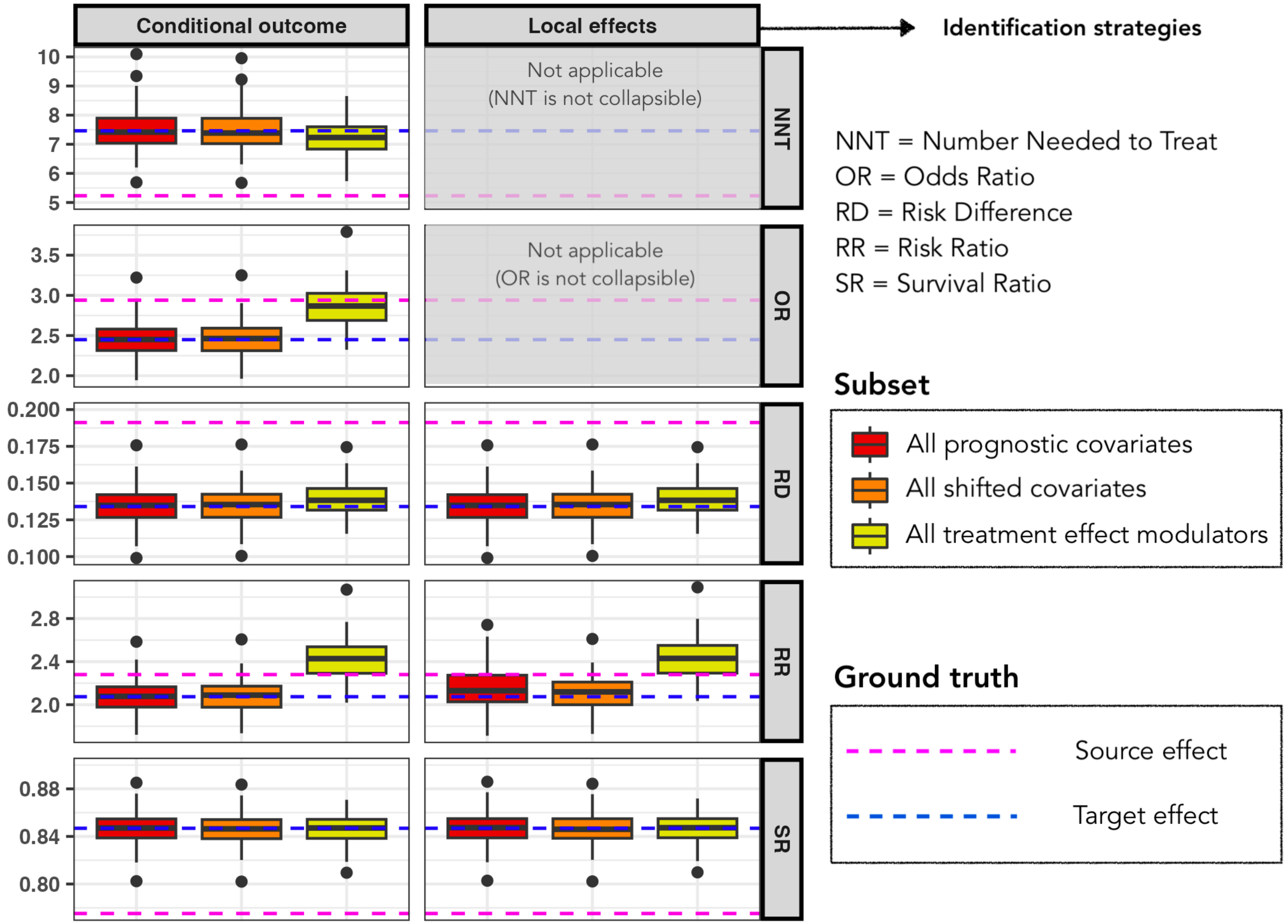
- Probability to die varies
  - Stressed people can die from a heart attack
  - Executioner more merciful when facing women

$$P[Y = 1 \mid X] = b(X_{1\to 3}) + (1 - b(X_{1\to 3})\, m(X_{2\to 3})$$

X1 : lifestyle general level

X2 : stress

X3 : gender (not shifted)



— Local SR can be generalised using only stress. All others measures requires lifestyle and stress.

Conclusion

1. A collapsible measure is needed to generalize local effects,

2. Some measures disentangle the baseline risk from the effect — and this depends on the outcome nature
   - If Y is continuous — Risk Difference
   - If Y is binary — Risk Ratio or Survival Ratio depending on the direction of effect

3. Generalization can be done under different assumptions, with
   - more or less baseline covariates
   - access to Y(0) in the target population or not

ArXiv

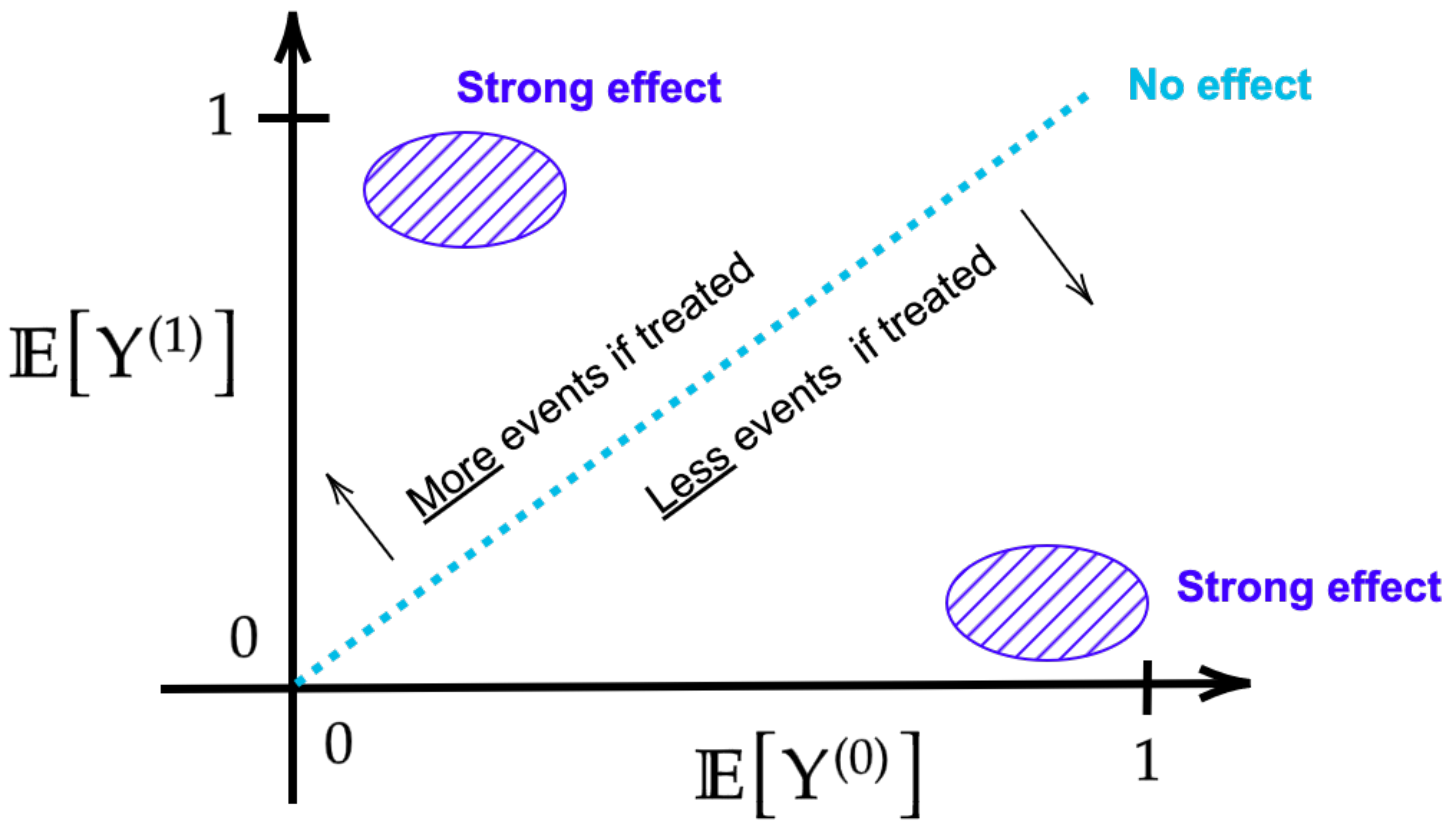
- Many thanks to Anders Huitfeldt, whose work inspired us!
- See Andrew Gelman's blog. Feel free to react!

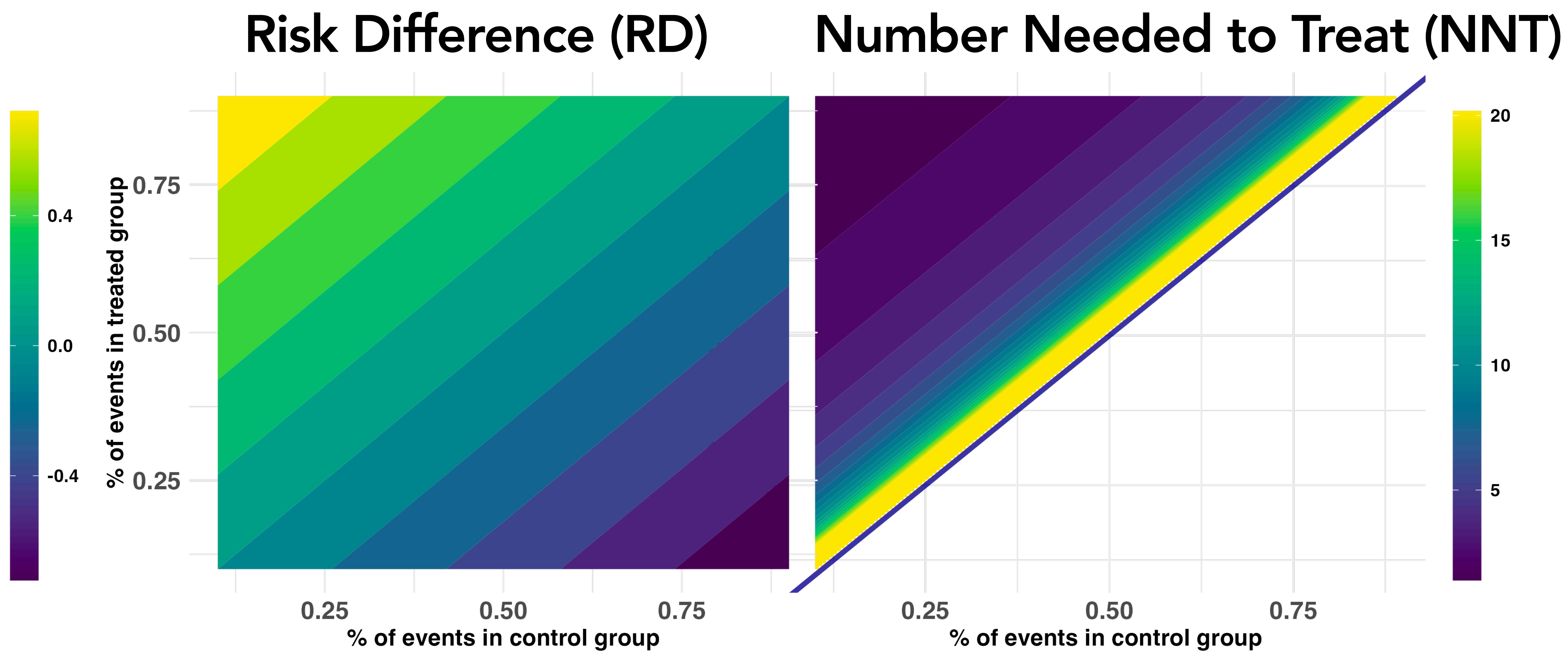
Thank you for listening!
Any questions?

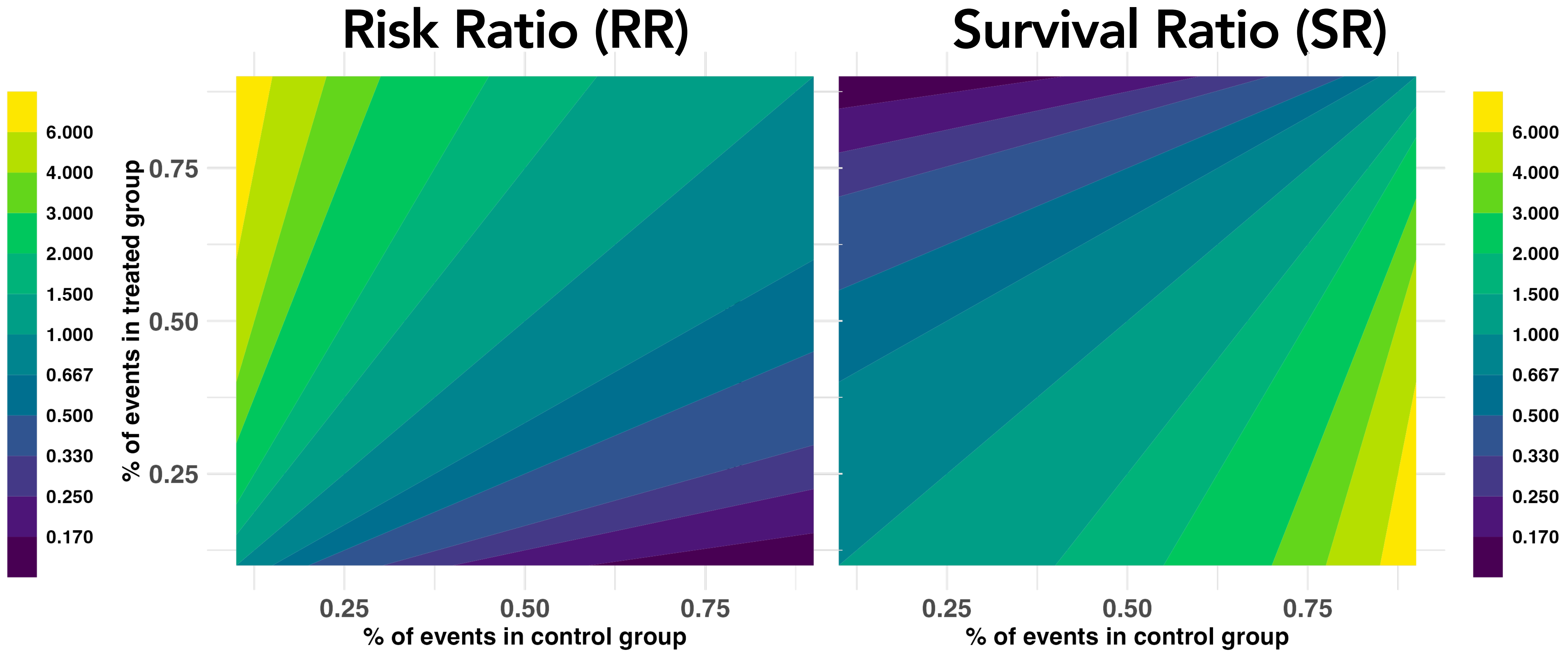
@BenedicteColnet

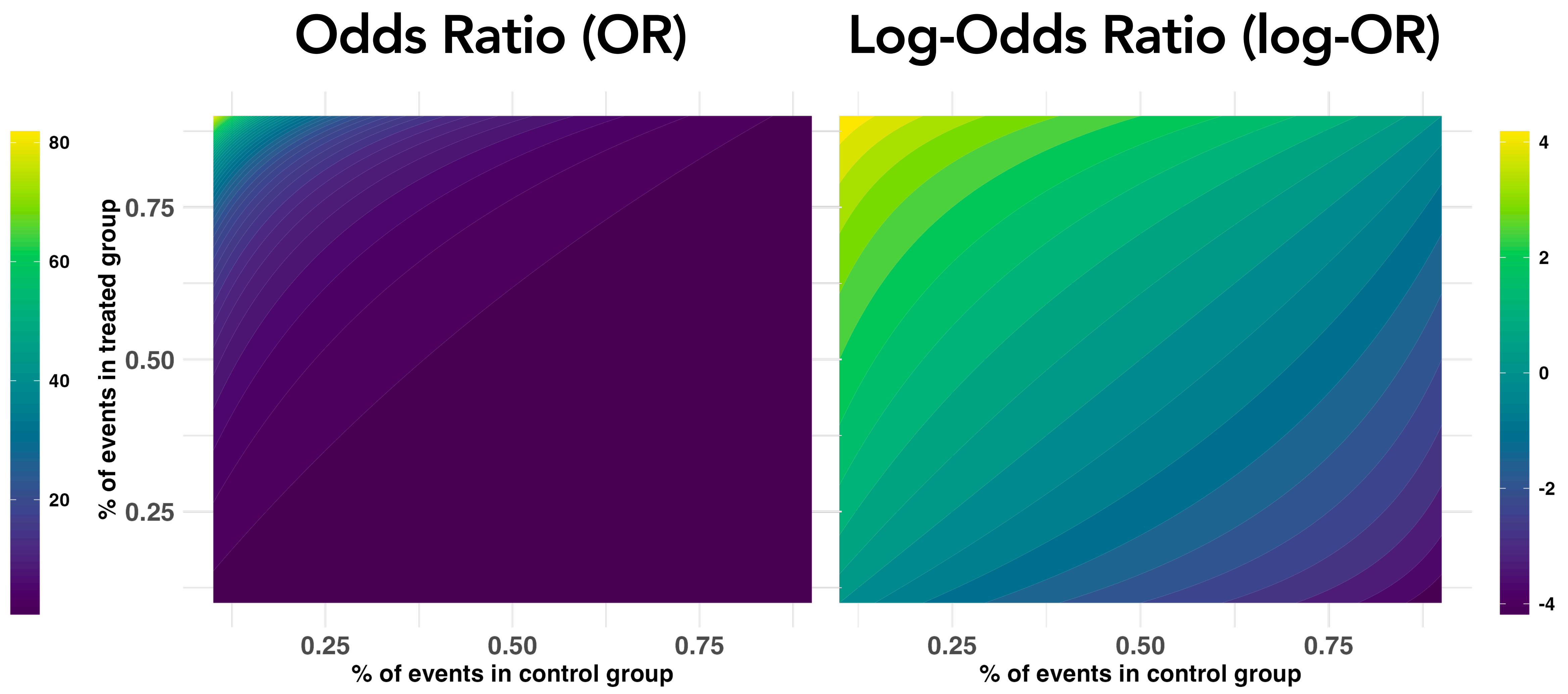
Ranges of effects

How to read plots







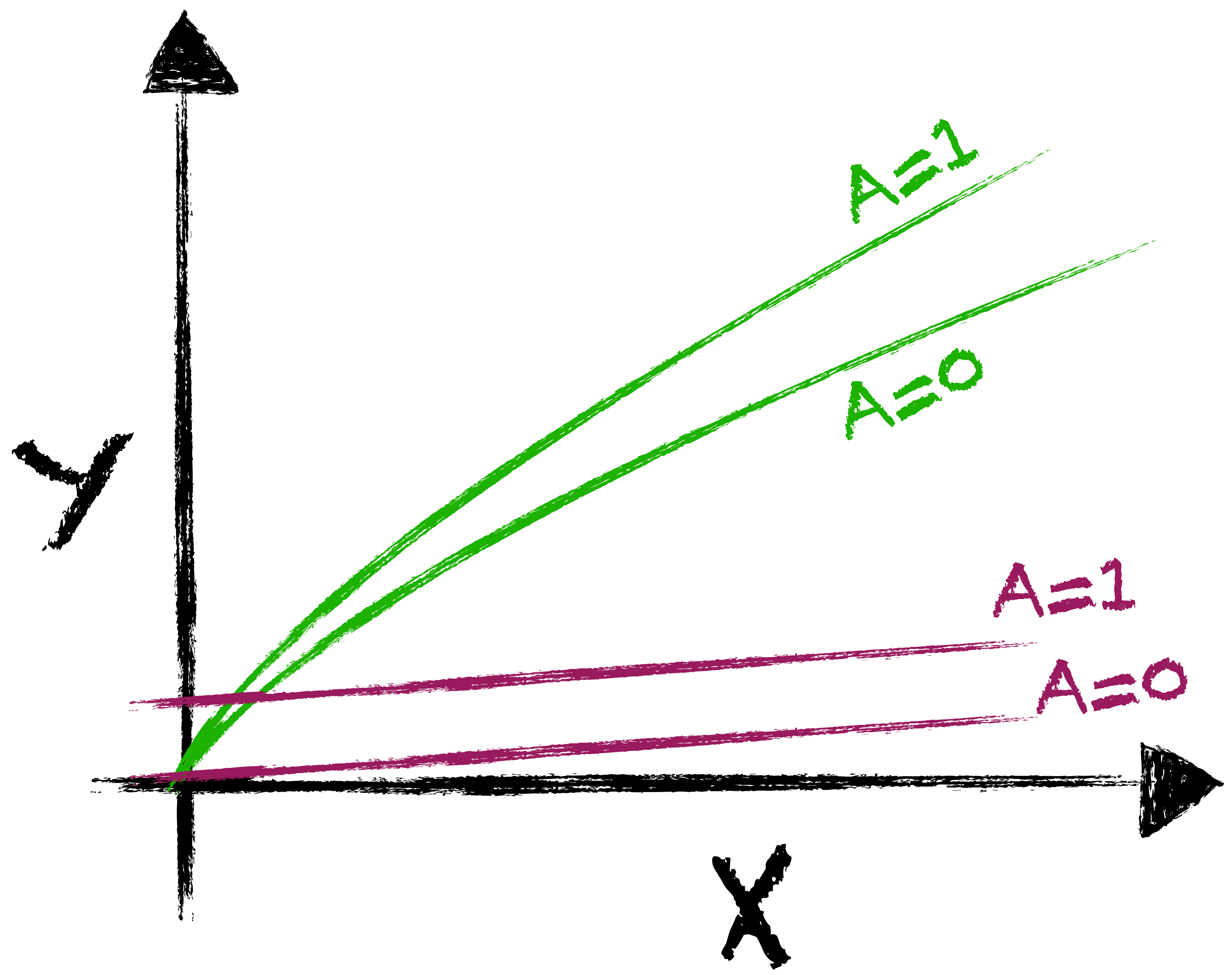


Common properties discussed

How the effect changes on sub-groups

Homogeneity $\forall x_1, x_2 \in \mathbb{X}, \quad \tau(x_1) = \tau(x_2) = \tau$

Heterogeneity $\exists x_1, x_2 \in \mathbb{X}, \quad \tau(x_1) \neq \tau(x_2)$



⚠️ *No non-zero effect can be homogeneous on all metrics*

How the effect changes with labelling

e.g. Odds Ratio is symmetric, while Risk Ratio is not