Risk ratio, odds ratio, risk difference...

Which causal measure is easier to generalize?

Bénédicte Colnet, Ph.D. student at Inria (Soda & PreMeDICaL teams)
Department of Statistics, University of Oxford, May 19th

Julie Josse
Missing values & causal inference

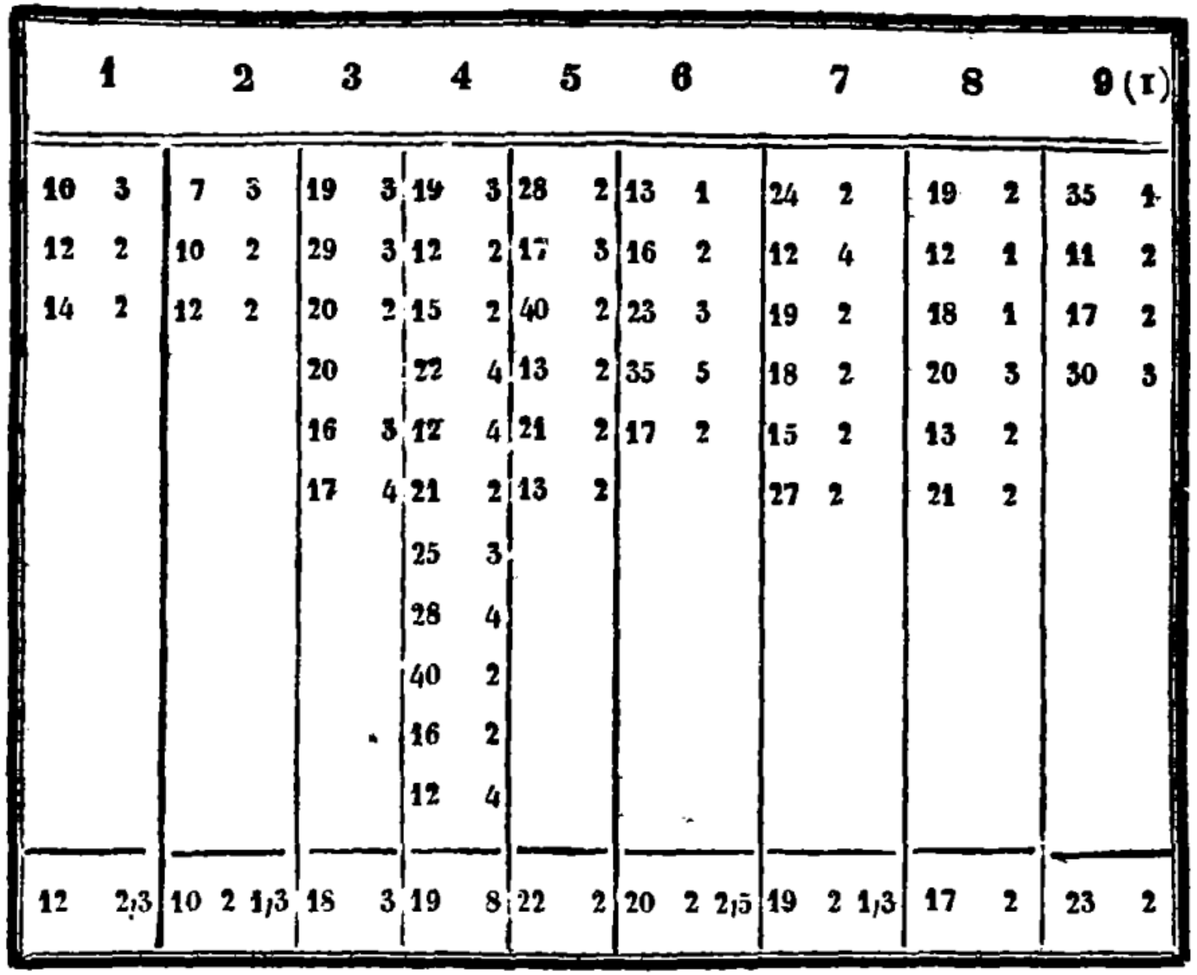
Gaël Varoquaux
ML & co-founder of scikit-learn

Erwan Scornet
Random forest & missing values

ArXiv

Evidence based medicine

The promise of big data



1		2		3		4		5		6		7		8		9 (1)	
16	3	7	3	19	3	19	3	28	2	13	1	24	2	19	2	35	1
12	2	10	2	29	3	12	2	17	3	16	2	12	4	12	1	11	2
14	2	12	2	20	2	15	2	40	2	23	3	19	2	18	1	17	2
				20		22	4	13	2	35	5	18	2	20	3	30	3
				16	3	17	4	21	2	17	2	15	2	13	2		
				17	4	21	2	13	2			27	2	21	2		
						25	3										
						28	4										
						40	2										
						16	2										
						12	4										
12	2,3	10	2 1/3	18	3	19	8	22	2	20	2 2,5	19	2 1/3	17	2	23	2

Source: Pierre Charles Alexandre Louis's experiment on bloodletting (1835)
— Original research work is made available by the French National Library (BnF)

A brief history of modern medical evidence: the ever increasing role of data and statistics

James Lind's scorbut experiment

1747

A brief history of modern medical evidence: the ever increasing role of data and statistics

James Lind's scorbut experiment

William Farr — *General Register Office*

1747 — 1828 — 1837 — 1854 — 1912

P.C.A. Louis's experiments on bloodletting

John Snow's discovery on cholera

Janet Lane-Clayton pioneered the use of cohort studies and case control studies (benefit of breast feeding versus cow milk)

A brief history of modern medical evidence: the ever increasing role of data and statistics

James Lind's scorbut experiment

William Farr — *General Register Office*

1747 — 1828 — 1837 — 1854 — 1912 — 1948

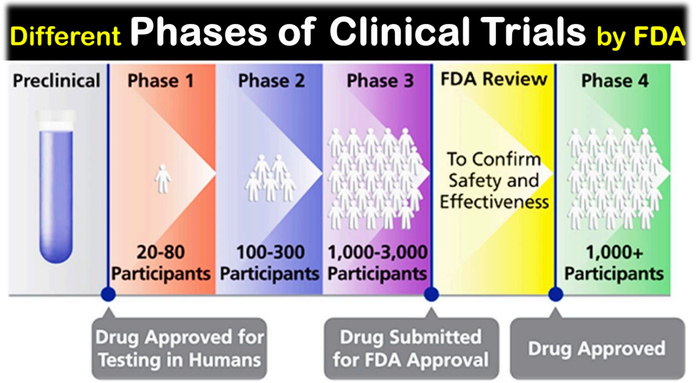
Streptomycin trial for pulmonary tuberculosis

So-called evidence based medicine's era

P.C.A. Louis's experiments on bloodletting

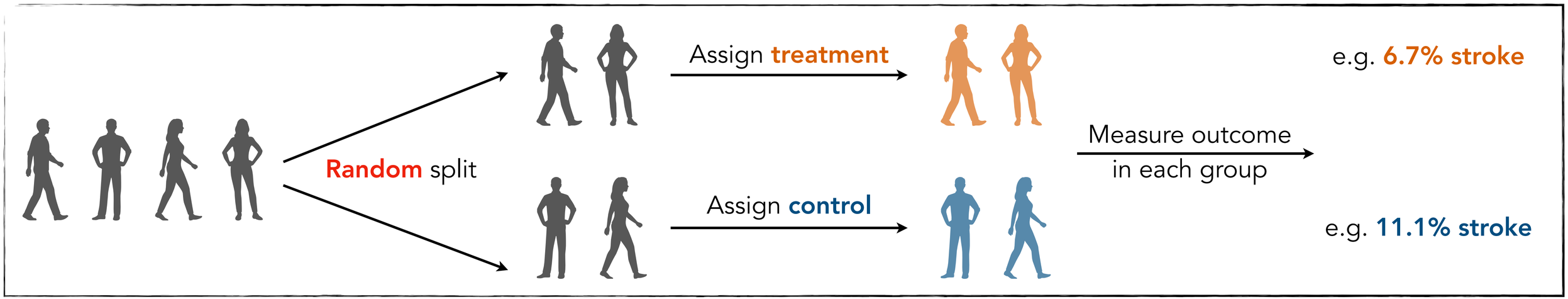
John Snow's discovery on cholera

Janet Lane-Clayton pioneered the use of cohort studies and case control studies (benefit of breast feeding versus cow milk)



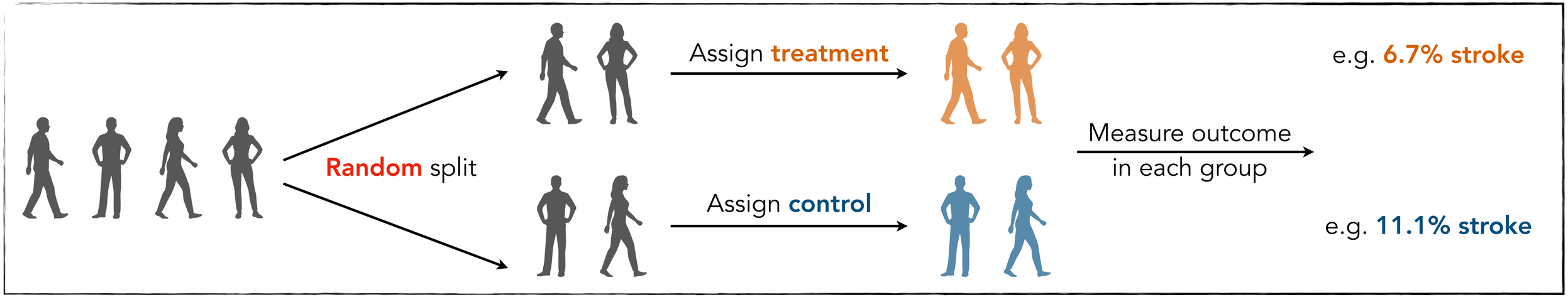
Randomized Controlled Trials (RCTs) as the current gold standard

Principle



Randomized Controlled Trials (RCTs) as the current gold standard

Principle

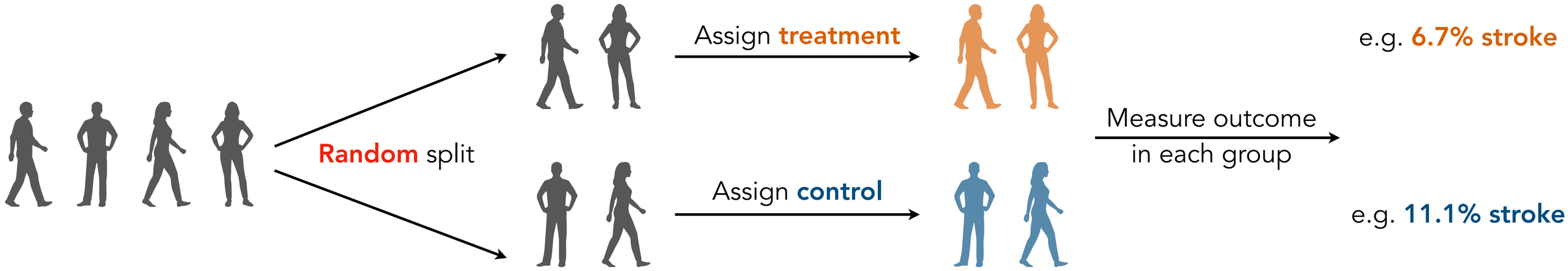


In practice : the CRASH-3 trial investigating Tranexamic Acid effect on brain injured related death

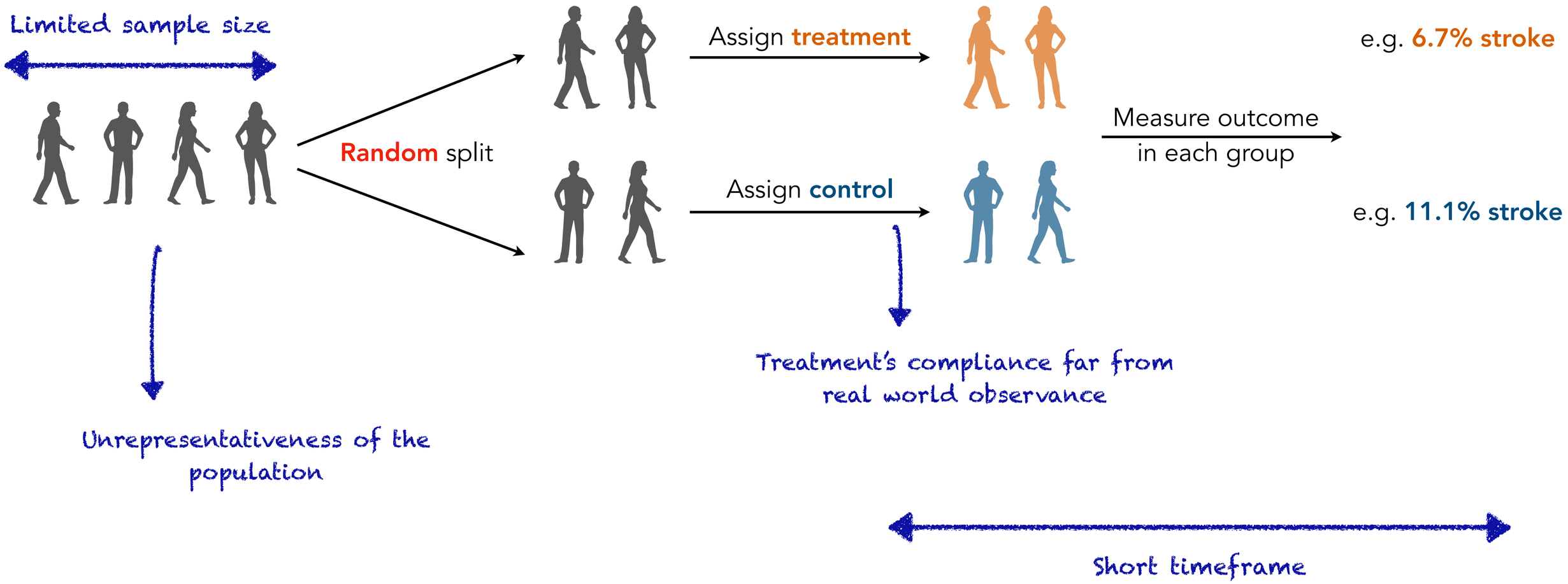
> **Results** Between July 20, 2012, and Jan 31, 2019, we randomly allocated 12 737 patients with TBI to receive tranexamic acid (6406 [50·3%] or placebo [6331 [49·7%], of whom 9202 (72·2%) patients were treated within 3 h of injury. Among patients treated within 3 h of injury, the risk of head injury-related death was 18·5% in the tranexamic acid group versus 19·8% in the placebo group (855 *vs* 892 events; risk ratio [RR] 0·94 [95% CI 0·86–1·02]).

Source: Screenshot from the Lancet (CRASH-3 main report)

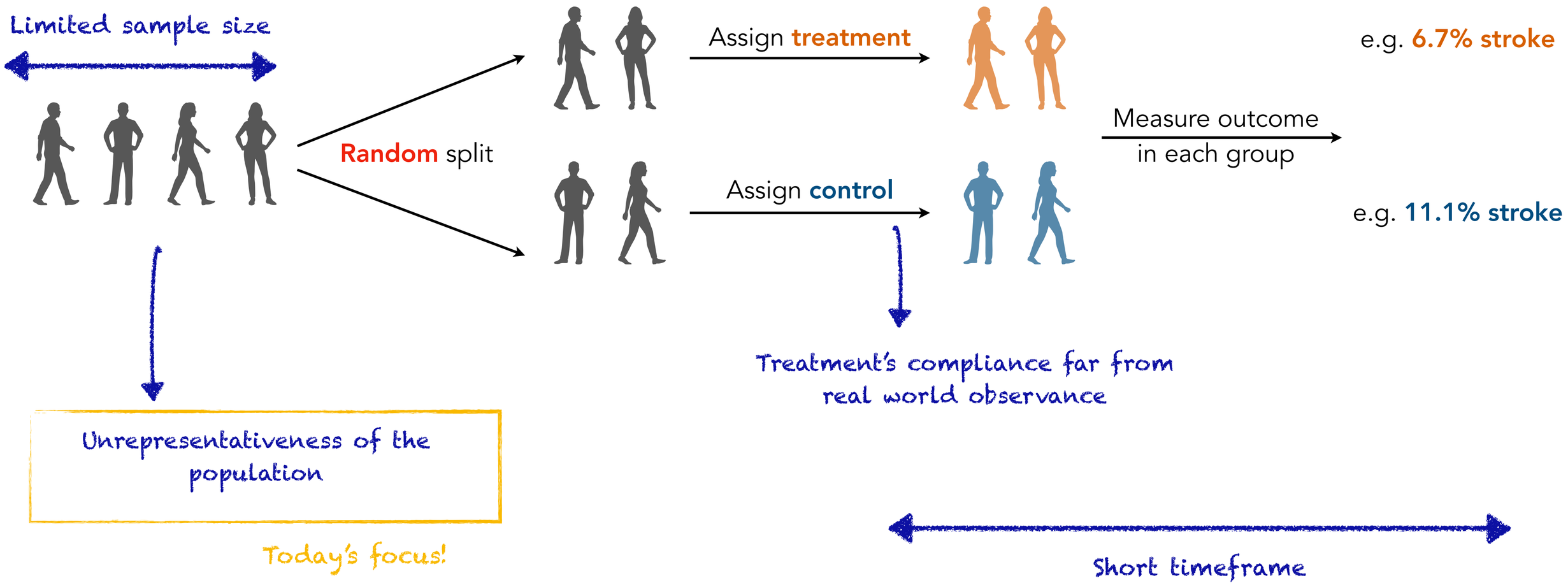
The limited scope of RCTs is increasingly under scrutiny



The limited scope of RCTs is increasingly under scrutiny



The limited scope of RCTs is increasingly under scrutiny



Our motivating example: generalization of CRASH-3 findings to the Traumabase

CRASH-3

- Multi-centric RCT with 9000 individuals
- Measured a positive effect on moderately injured patients

Traumabase

- Large national French cohort with 30000 individuals
- Could not conclude on a positive effect when adjusting on confounders

What would be the estimated effect of TXA if measured on the Traumabase's population?

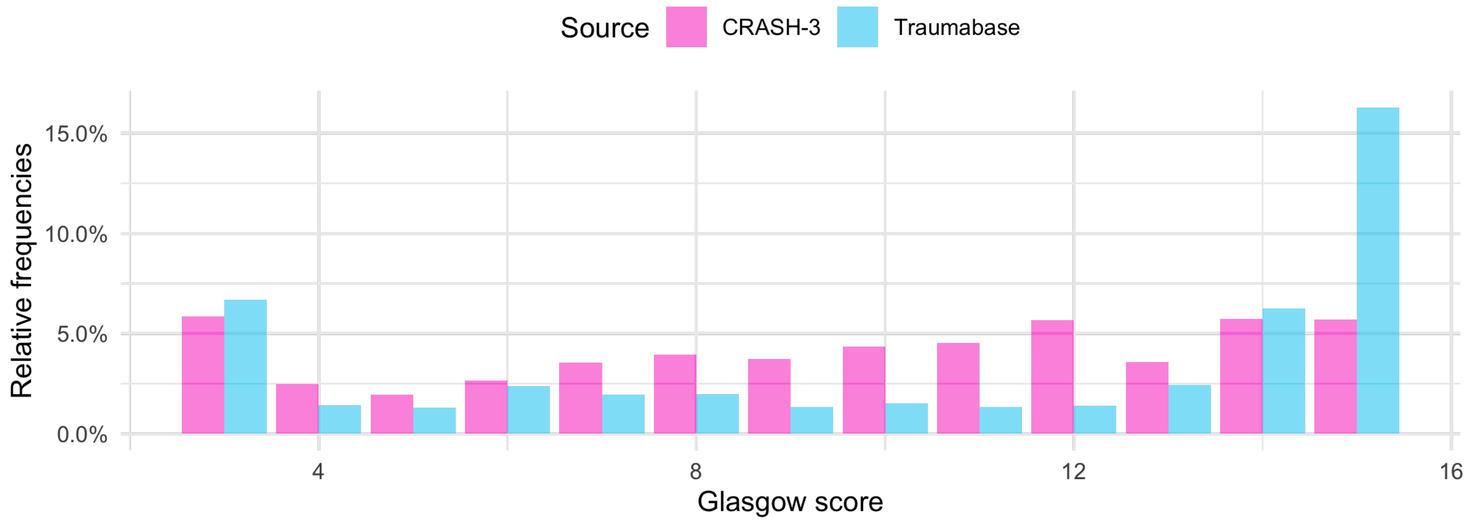
Our motivating example: generalization of CRASH-3 findings to the Traumabase

CRASH-3

- Multi-centric RCT with 9000 individuals
- Measured a positive effect on moderately injured patients

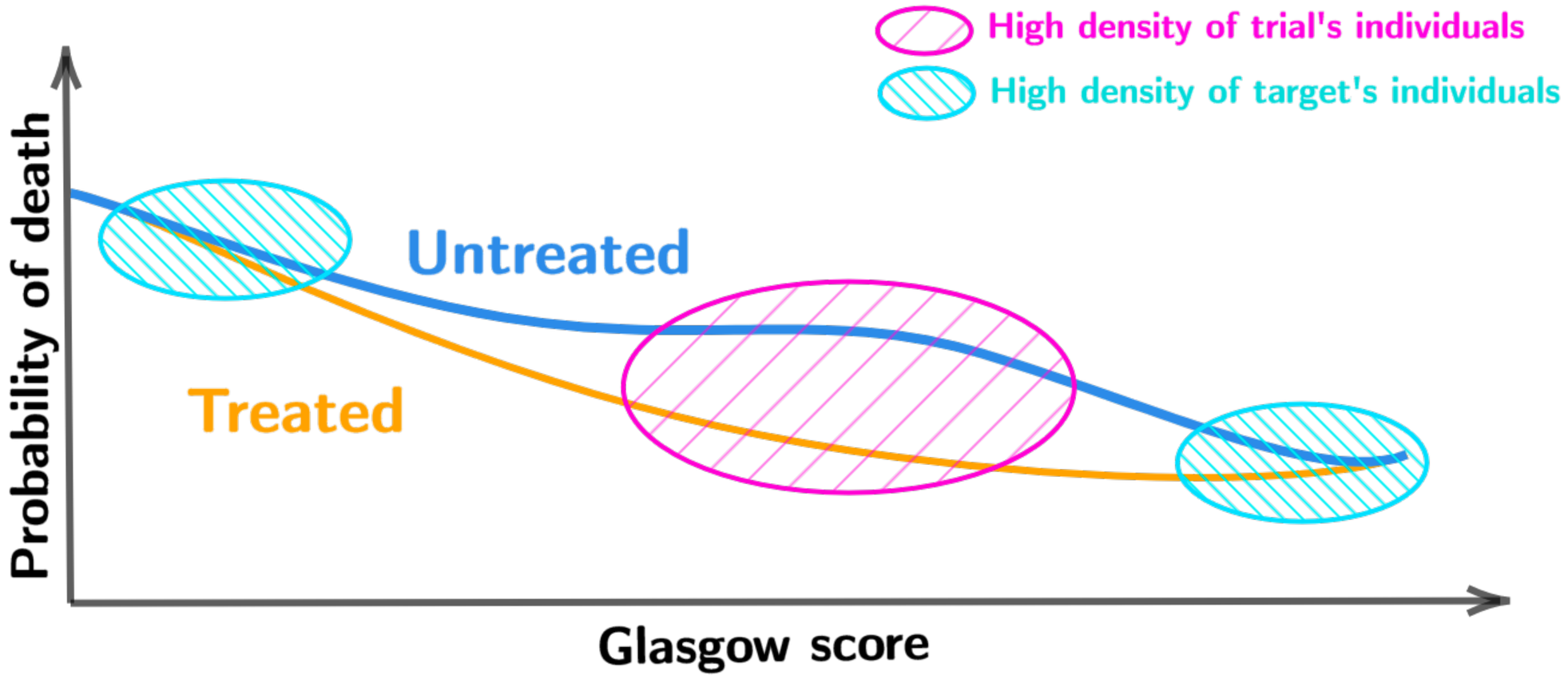
Traumabase

- Large national French cohort with 30000 individuals
- Could not conclude on a positive effect when adjusting on confounders



Can the result of a large international trial — assessing the efficacy of Tranexamic Acid (TXA) on brain-injured death (TBI) — be *generalized* to the French population?

What did you mean by *heterogeneity of treatment* effect?



Hypothetical drawing of the response model.
Glasgow score reflects the severity of the brain trauma, the lower the score the higher the trauma.

Toward formalization — the potential outcomes framework to encode causality

For each individual ***i***, consider each of the possible outcomes for treated $\mathbf{Y^{(1)}}$, and control $\mathbf{Y^{(0)}}$.

Toward formalization — the potential outcomes framework to encode causality

For each individual ***i***, consider each of the possible outcomes for treated $Y^{(1)}$, and control $Y^{(0)}$.

characteristics

binary treatment

X		A	Y(1)	Y(0)	Y
F	1	0	NA	3	3
M	2	0	NA	5	5
M	1	1	14	NA	14
F	3	0	NA	8	8
F	2	1	7	NA	7

Y is the observed outcome

Toward formalization — the potential outcomes framework to encode causality

For each individual ***i***, consider each of the possible outcomes for treated $Y^{(1)}$, and control $Y^{(0)}$.

characteristics

binary treatment

X		A	Y(1)	Y(0)	Y
F	1	0	NA	3	3
M	2	0	NA	5	5
M	1	1	14	NA	14
F	3	0	NA	8	8
F	2	1	7	NA	7

Y is the observed outcome

In a RCT, $$\frac{1}{n_1}\sum_{i=1}^{n} A_i Y_i \rightarrow \mathbb{E}\left[Y \mid A = 1\right] = \mathbb{E}\left[Y^{(1)}\right]$$

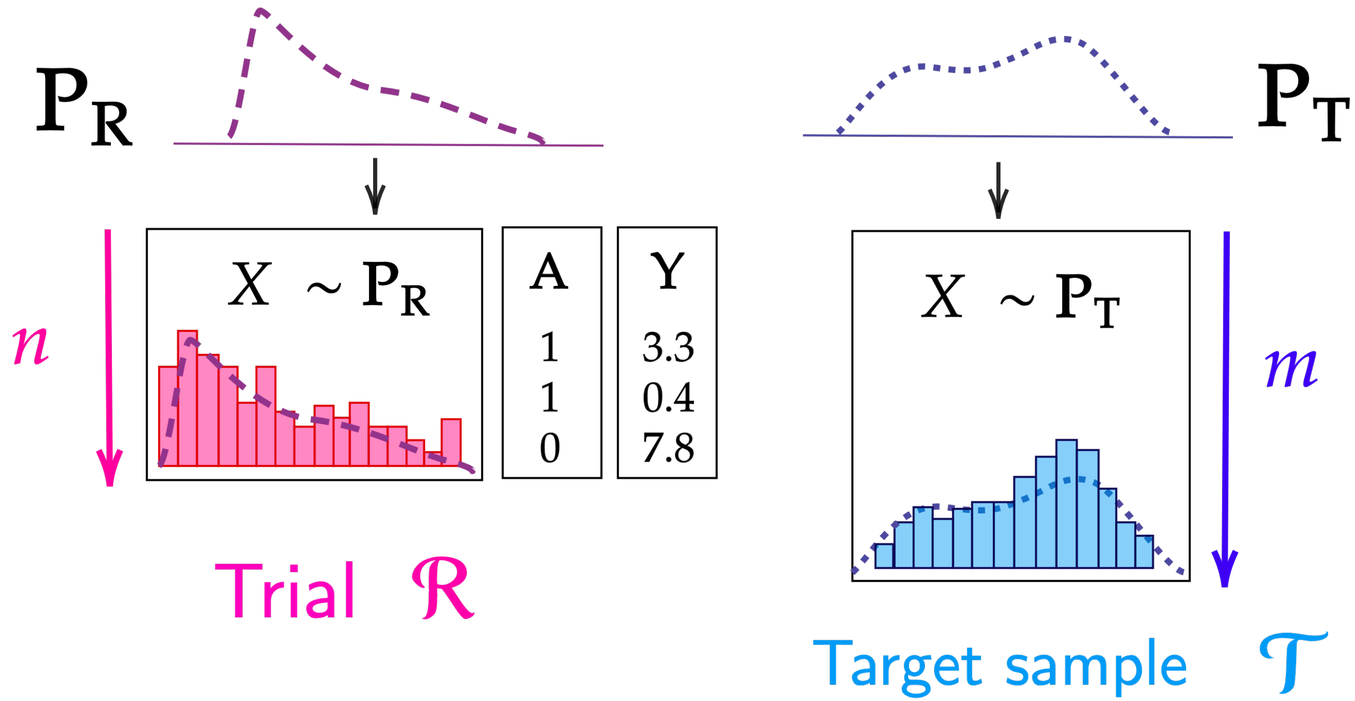
The potential outcomes framework for the generalization

Denoting,

- **A** the binary treatment
- **X** the covariates
- **Y** the observed outcome

We now consider,

- A **trial** of size **n** sampled from a population $p_R(X)$,
- A data set of size **m** sampled from $p_T(X)$ the **target** population of interest.



Generalizing clinical trial's findings

When estimation depends on two data sets

Recalling what is done on a classical clinical randomized trial

Horvitz-Thomson estimator

$$\hat{\tau}_{HT,n} = \frac{1}{n} \sum_{i \in Trial} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1 - A_i)}{1 - \pi} \right)$$

Probability to receive treatment, usually 0.5

Recalling what is done on a classical clinical randomized trial

Horvitz-Thomson estimator

$$\hat{\tau}_{HT,n} = \frac{1}{n} \sum_{i \in Trial} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1 - A_i)}{1 - \pi} \right)$$

Probability to receive treatment, usually 0.5

Properties

$$\mathbb{E}\left[\hat{\tau}_{HT,n}\right] = \tau$$

Unbiased

$$n \operatorname{Var}\left[\hat{\tau}_{HT,n}\right] = \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{\pi} + \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{1 - \pi} - \tau^2 := V_{HT}$$

Finite sample variance

Enriching the trial data with the target sample data

IPSW

Depends on n and m !

$$\hat{\tau}_{IPSW,n,m} = \frac{1}{n} \sum_{i \in Trial} \frac{\hat{p}_{T,m}(X_i)}{\hat{p}_{R,n}(X_i)} \left(\frac{Y_i A_i}{\pi} - \frac{Y_i(1 - A_i)}{1 - \pi} \right)$$

Same as single RCT

Wished properties?

$$\mathbb{E}\left[\hat{\tau}_{IPSW,n}\right] = \tau_T$$

Unbiased

$$n \operatorname{Var}\left[\hat{\tau}_{IPSW,n,m}\right] = ?$$

Generalization's *causal* assumptions

Transportability assumption

$$\forall x \in \mathbb{X}, \quad \mathbb{P}_R(Y^{(1)} - Y^{(0)} \mid X = x) = \mathbb{P}_T(Y^{(1)} - Y^{(0)} \mid X = x)$$

→ Needed covariates are shifted treatment effect modifiers

Positivity assumption

$$\text{supp}(P_T(X)) \subset \text{supp}(P_R(X))$$

→ Each individuals in the target population has to be represented in the trial.

Our contributions

Assumption: assume **X** is composed of categorical covariates — e.g. smoking status, gender, …

$$\hat{p}_{R,n}(x) := \frac{1}{n} \sum_{i \in \mathbb{R}} 1_{X_i = x}$$

Our contributions

Assumption: assume **X** is composed of categorical covariates — e.g. smoking status, gender, …

$$\hat{p}_{R,n}(x) := \frac{1}{n} \sum_{i \in \mathbb{R}} 1_{X_i = x}$$

Asymptotic results for IPSW estimator

Letting $\lim_{n,m \to \infty} m/n = \lambda \in [0, \infty]$,

$$\lim_{n,m \to \infty} \min(n, m) \mathrm{Var}\left[\hat{\tau}_{n,m}\right] = \min(1, \lambda) \left(\frac{\mathrm{Var}\left[\tau(X)\right]}{\lambda} + V_{so} \right)$$

Variance depends on two data samples sizes!

Impact of additional covariates: for the worse?

- Covariates needed: shifted covariates and treatment effect modifiers

- One may be tempted to add many covariates

- But what happen if adding shifted covariates that are not modulating treatment effect? e.g. gender?

Impact of additional covariates: for the worse?

- Covariates needed: shifted covariates and treatment effect modifiers

- One may be tempted to add many covariates

- But what happen if adding shifted covariates that are not modulating treatment effect? e.g. gender?

$$\lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}_{T,n,m}(X,V)\right] = \left(\sum_{v\in\mathscr{V}} \frac{p_T(v)^2}{p_R(v)}\right) \lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}_{T,n,m}(X)\right]$$

Inflation x Variance without gender

Impact of additional covariates: for the worse?

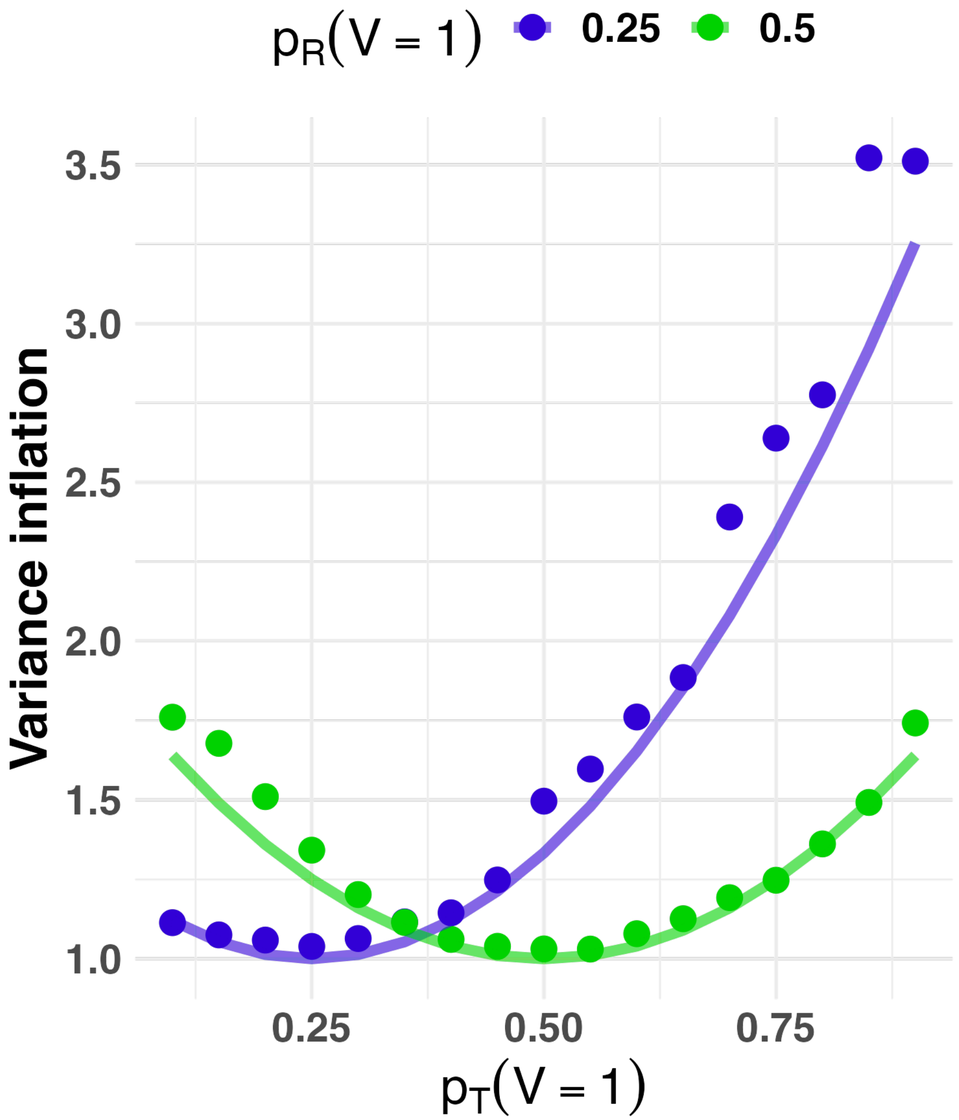
- Covariates needed: shifted covariates and treatment effect modifiers

- One may be tempted to add many covariates

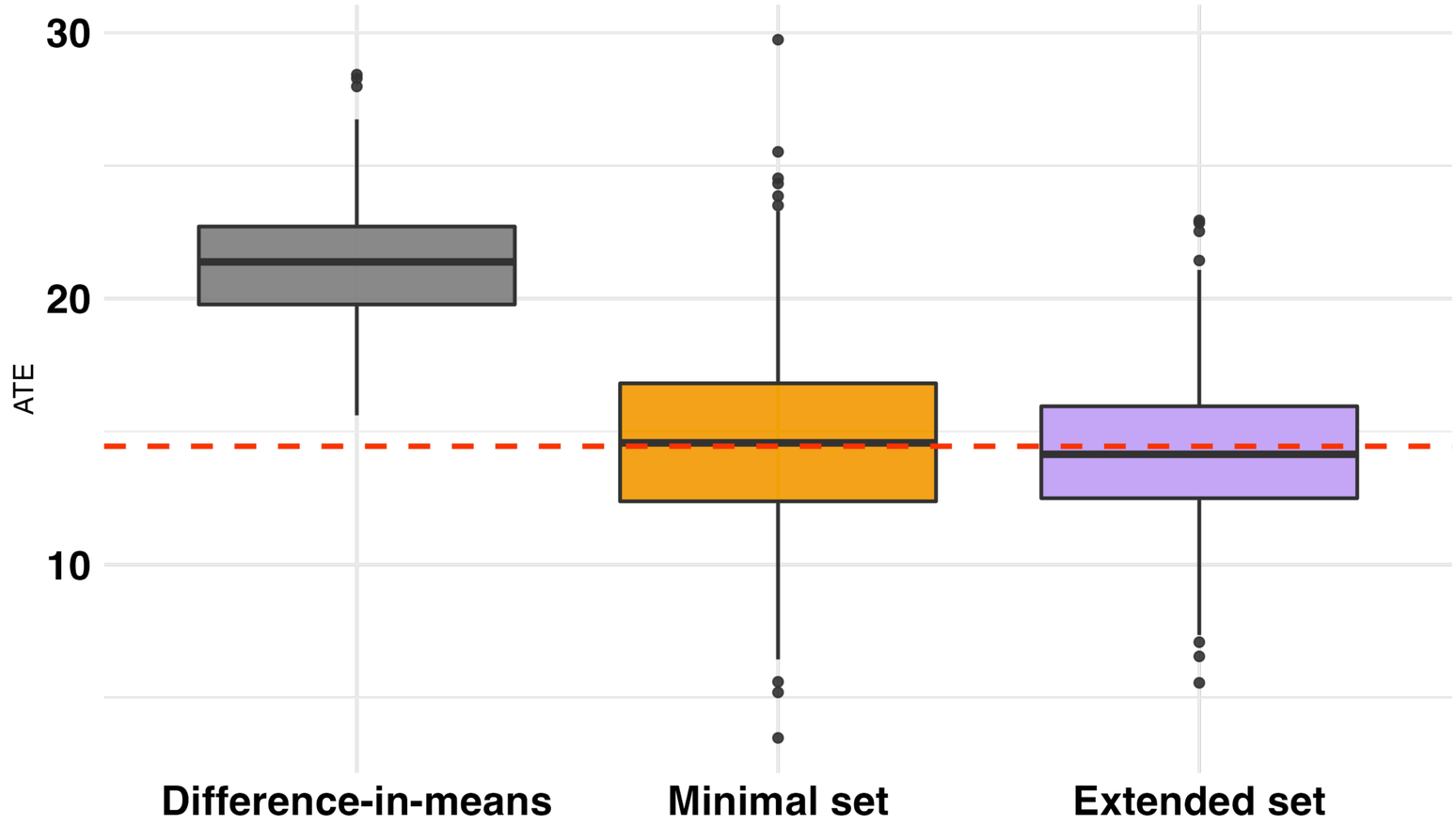
- But what happen if adding shifted covariates that are not modulating treatment effect? e.g. gender?

$$\lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}_{T,n,m}(X,V)\right] = \left(\sum_{v\in\mathscr{V}} \frac{p_T(v)^2}{p_R(v)}\right) \lim_{n\to\infty} n \operatorname{Var}_R\left[\hat{\tau}_{T,n,m}(X)\right]$$

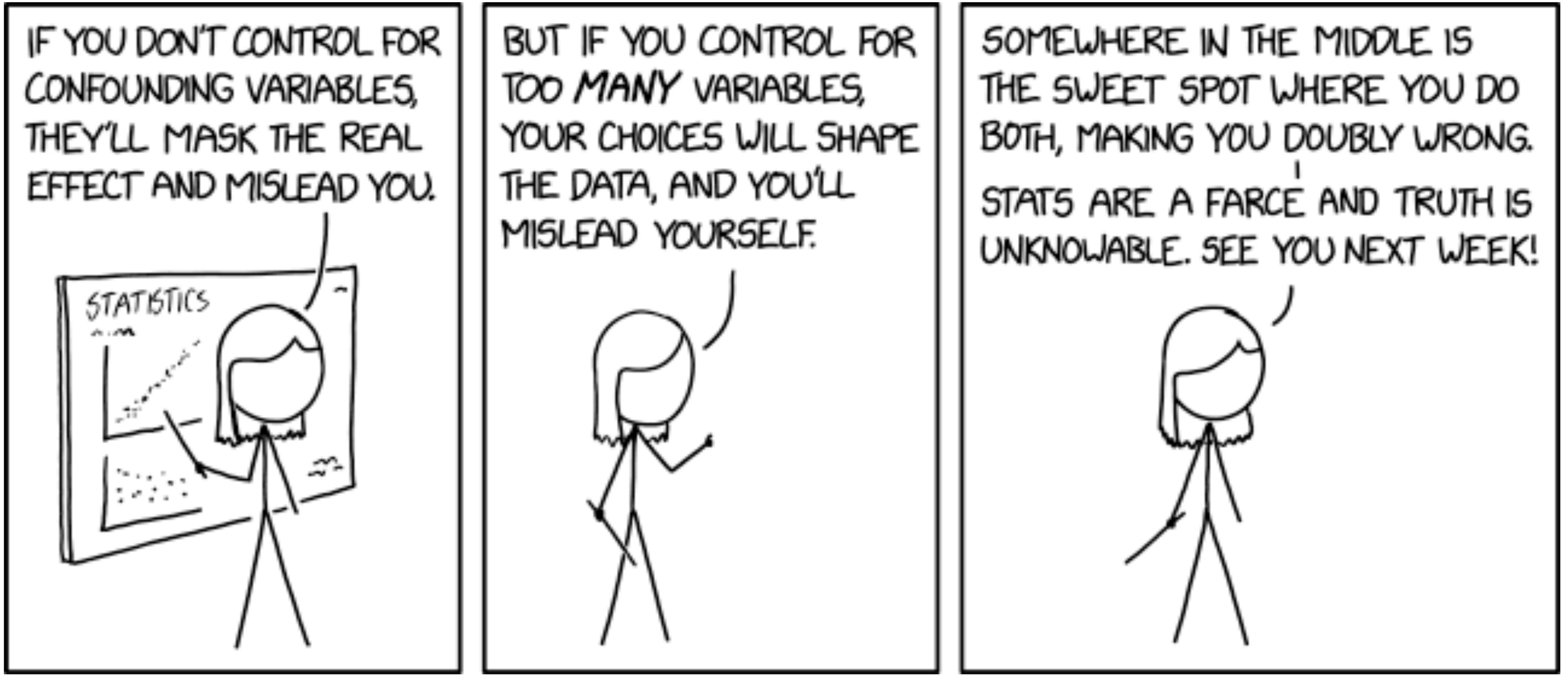
Inflation x Variance without gender



Impact of additional covariates: for the better?



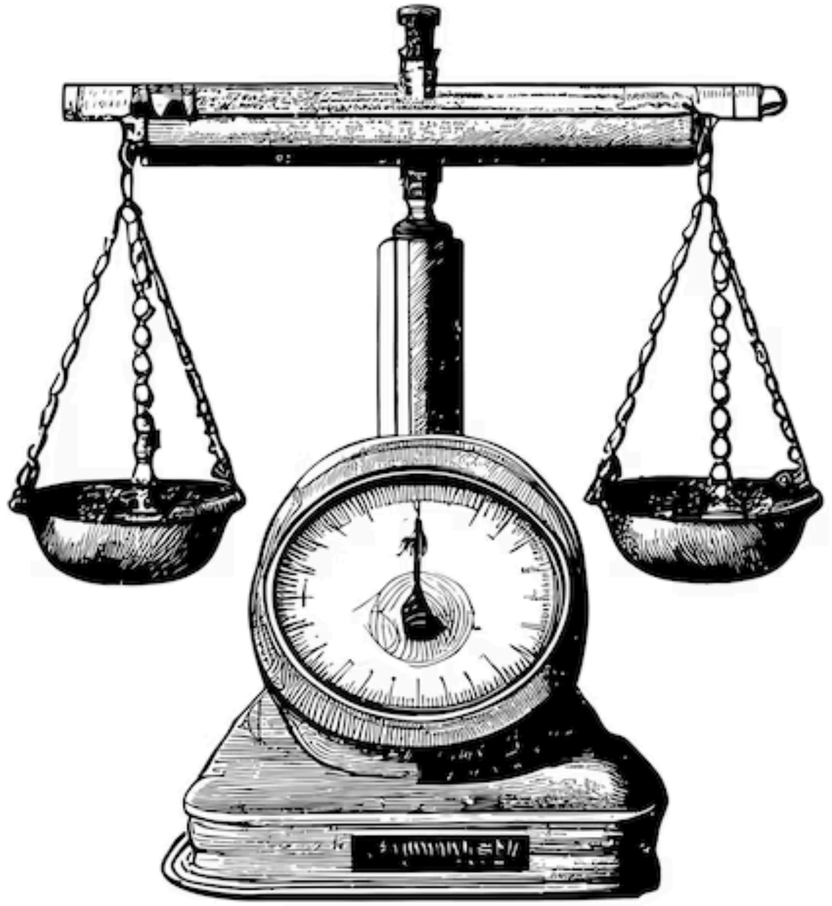
Adding a non-shifted, but treatment effect modifiers covariate, in the adjustment set **improves** precision



Source: xkcd.com

Risk ratio, odds ratio, risk difference

Which causal measure is easier to generalize?



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

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

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

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

Marginal effects τ

Conditional effects $\tau(x)$

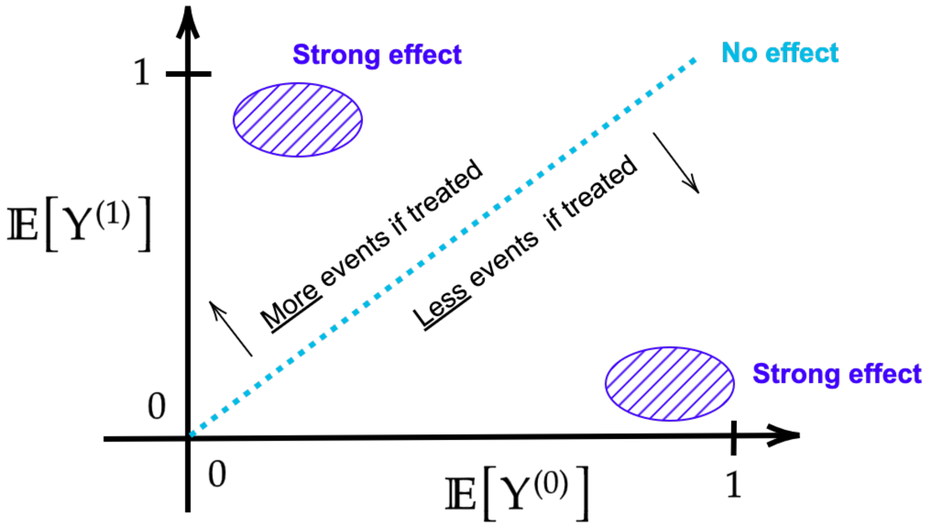
Computed from Cook & Sackett (1995)

👩‍⚕️ *``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.''*

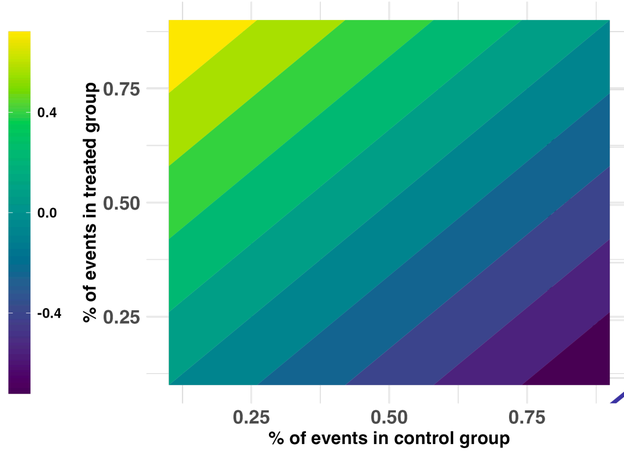
— leading to different impressions and heterogeneity patterns

Ranges of effects

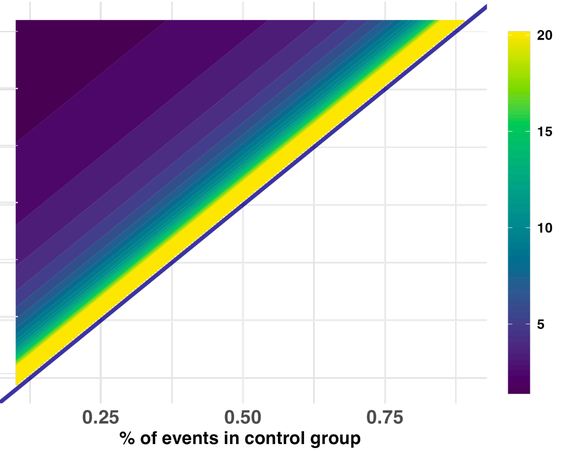
How to read plots



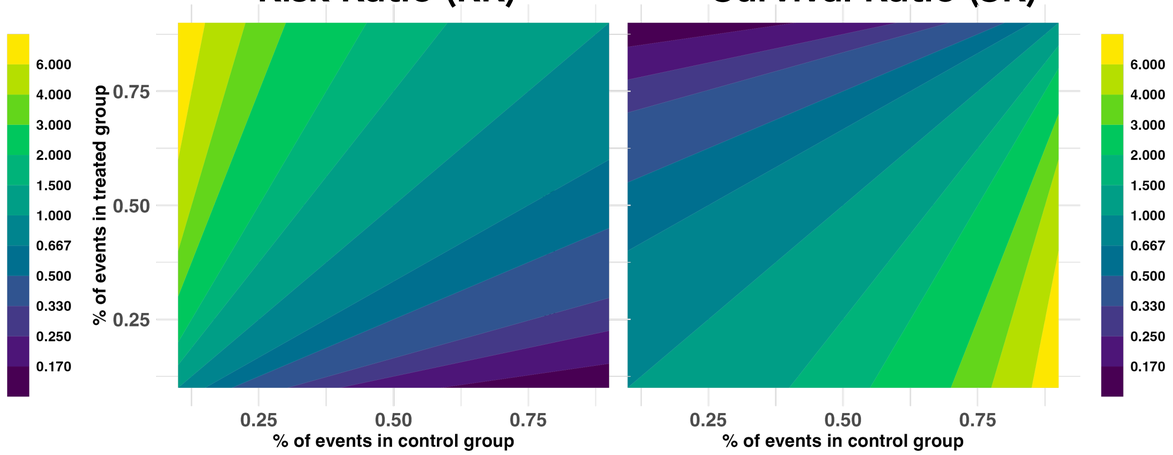
Risk Difference (RD)



Number Needed to Treat (NNT)

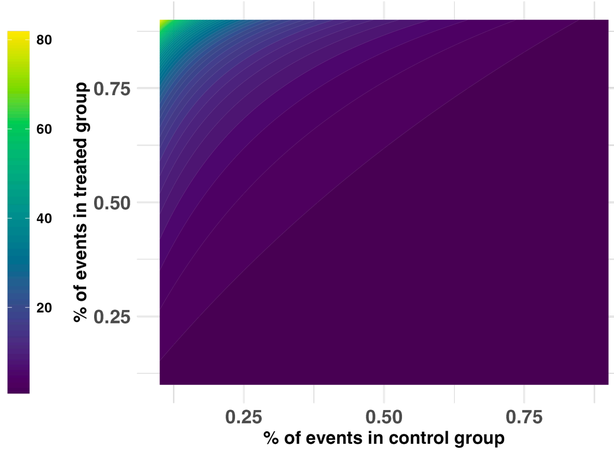


Risk Ratio (RR)

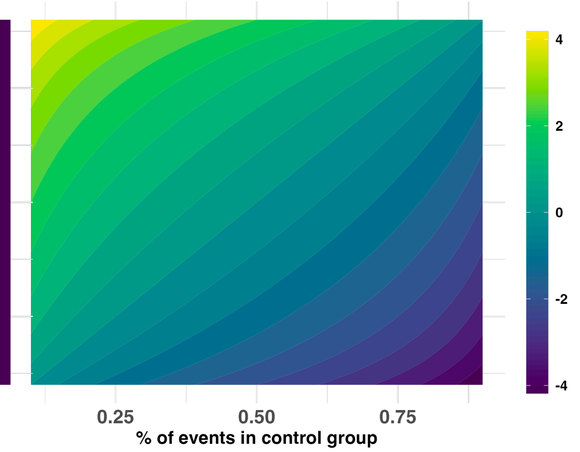


Survival Ratio (SR)

Odds Ratio (OR)



Log-Odds Ratio (log-OR)



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 $\tau \in \left[\min_x(\tau(x)), \max_x(\tau(x))\right]$

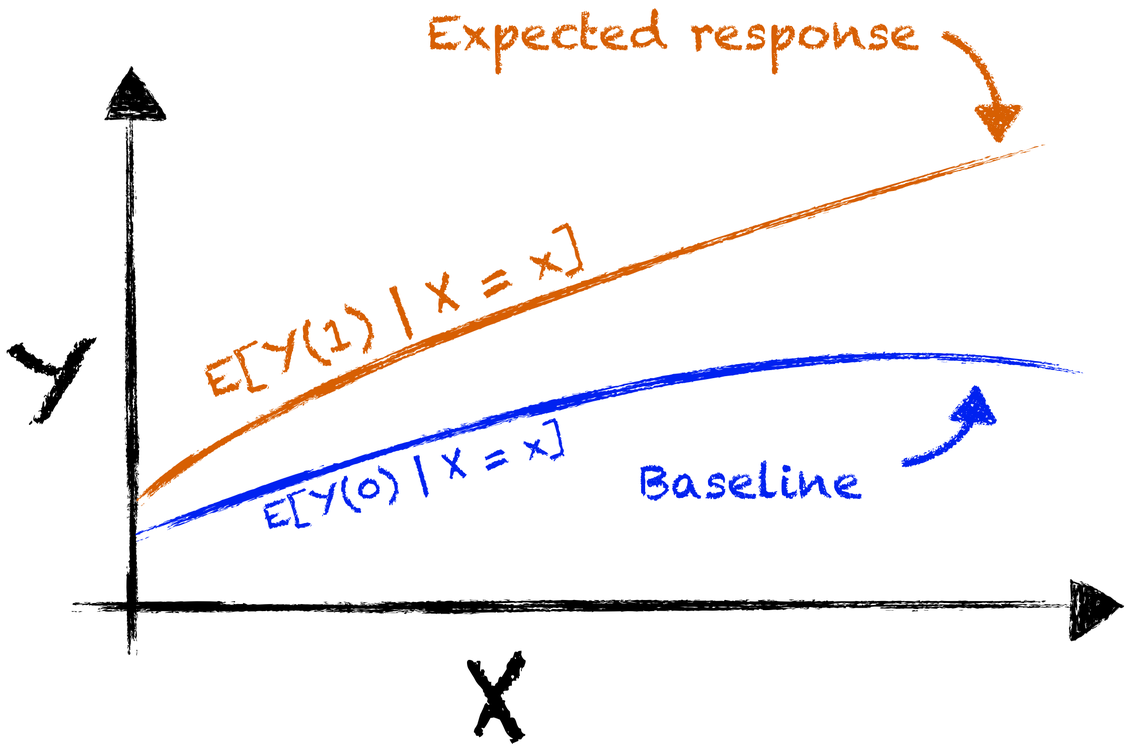
e.g RR is collapsible, with

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

Measure	Collapsible	Logic-respecting
Risk Difference (RD)	Yes	Yes
Number Neeeded 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

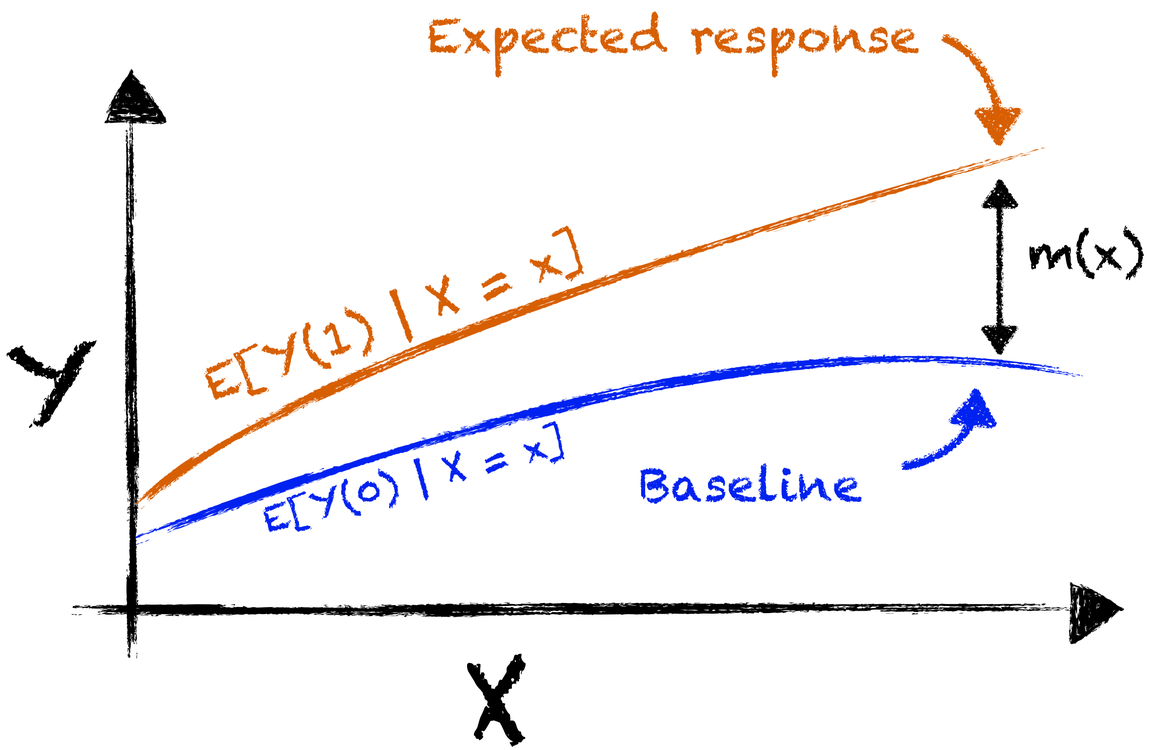
For Y <u>continuous</u>,



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

Through the lens of non parametric generative models

For Y <u>continuous</u>,



(*) 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

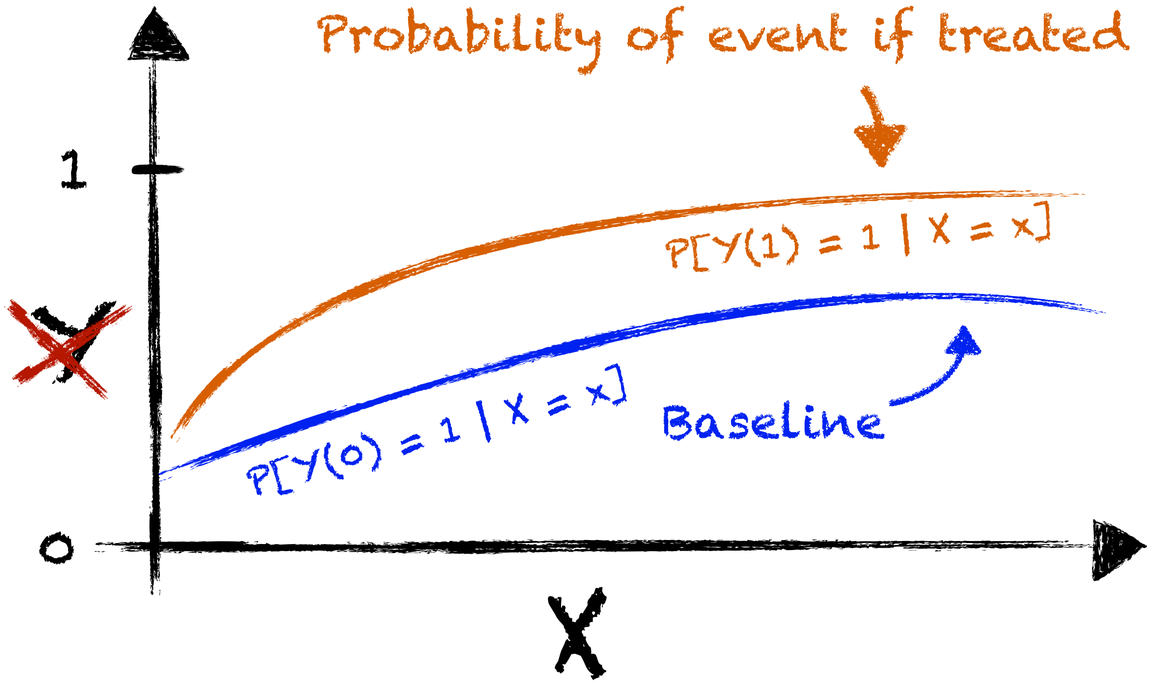
Linking generative functions with measures

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

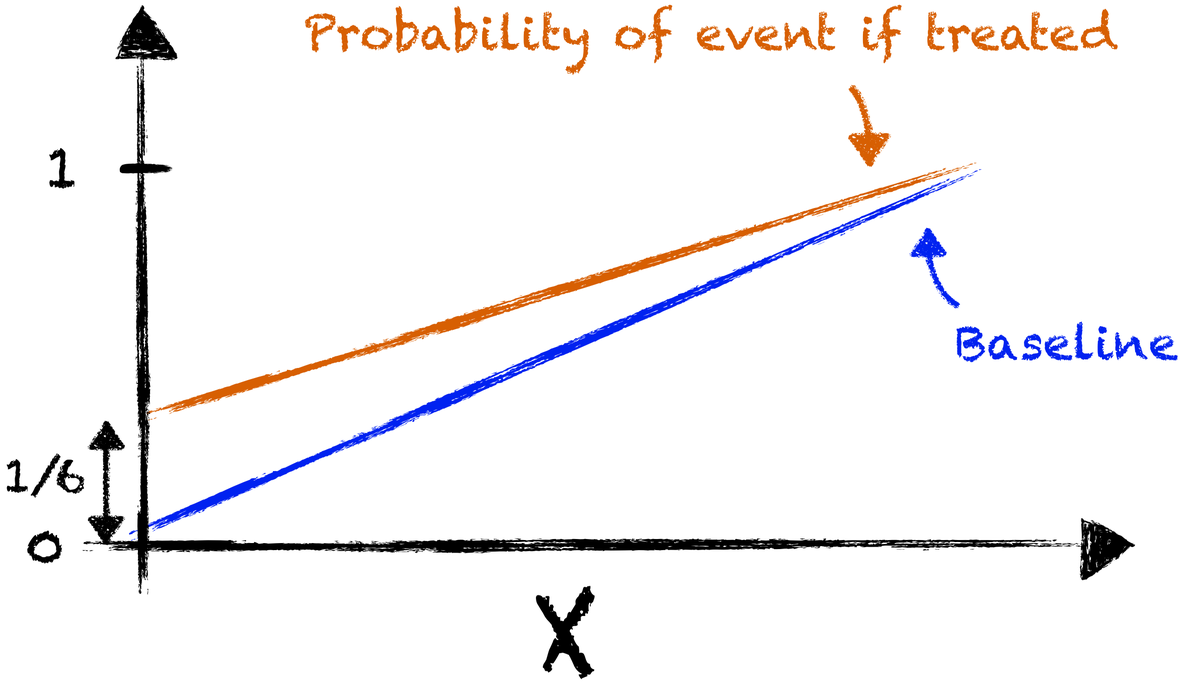
There exist two functions b(.) and m(.) such that,

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

The example of the Russian roulette

Harmful

For Y <u>binary</u>,

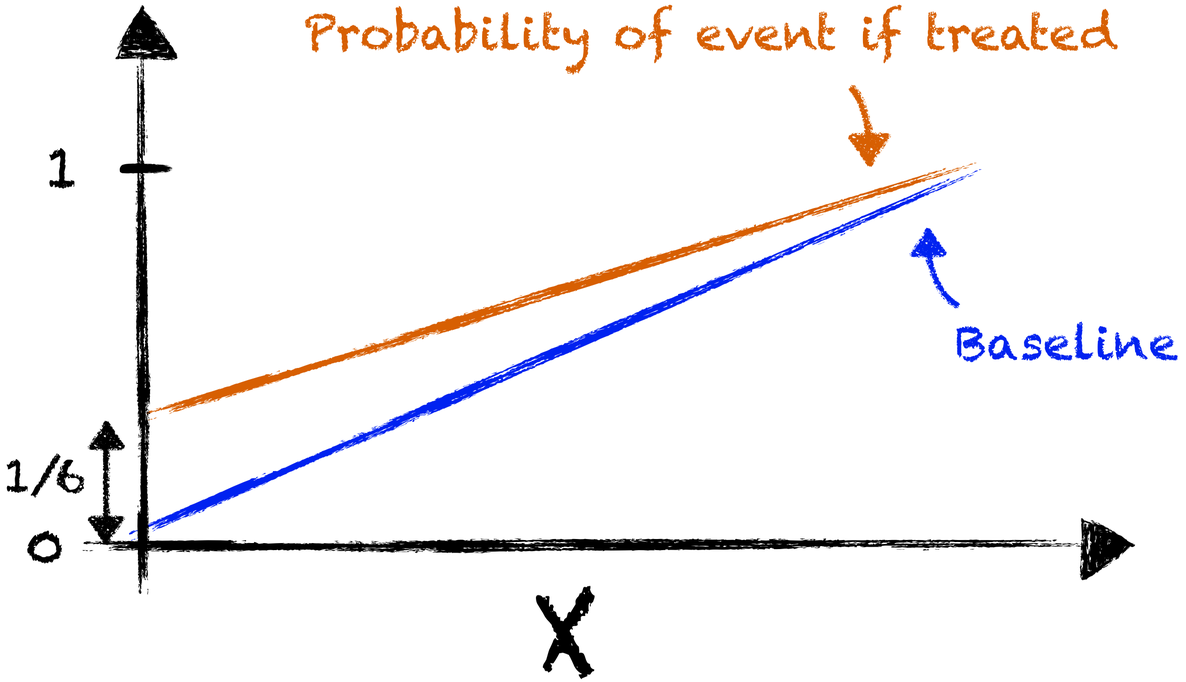


Example from Anders Huitfeldt, further used in Cinelli & Pearl (2020)

The example of the Russian roulette

Harmful

For Y <u>binary</u>,



Lemma

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

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

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,

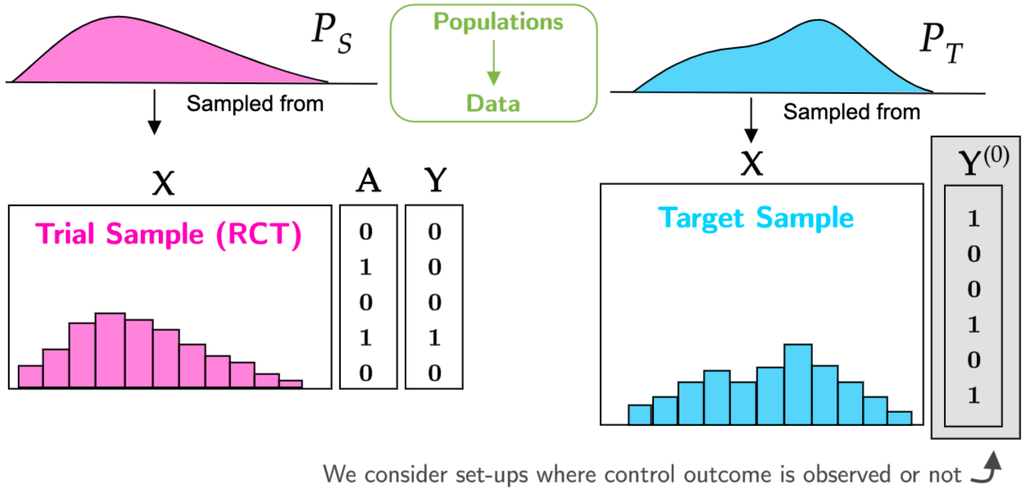
$$m_g(x) := \mathbb{P}\left[Y^{(1)} = 0 \mid Y^{(0)} = 1, X = x\right] \quad \textbf{and} \quad m_b(x) := \mathbb{P}\left[Y^{(1)} = 1 \mid Y^{(0)} = 0, X = x\right],$$

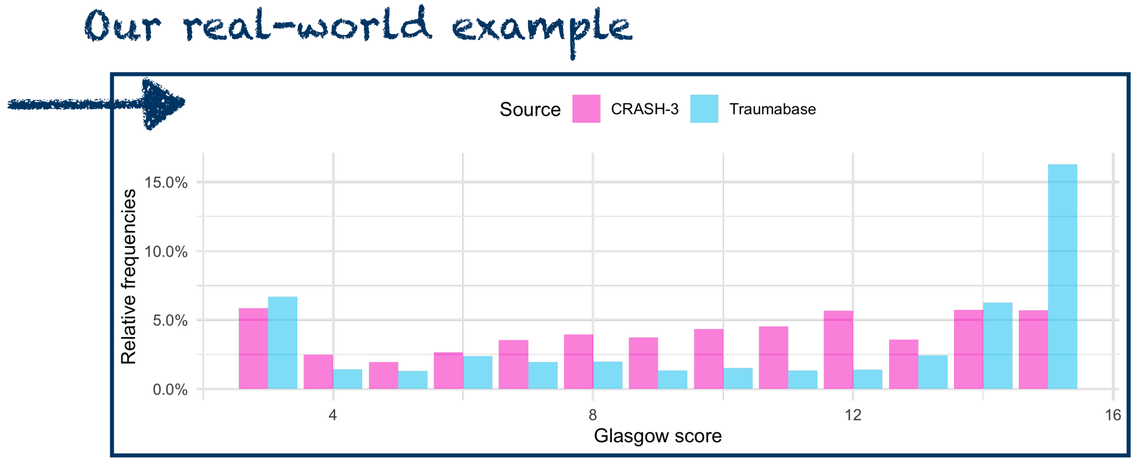
allows to have,

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

Generalizability

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

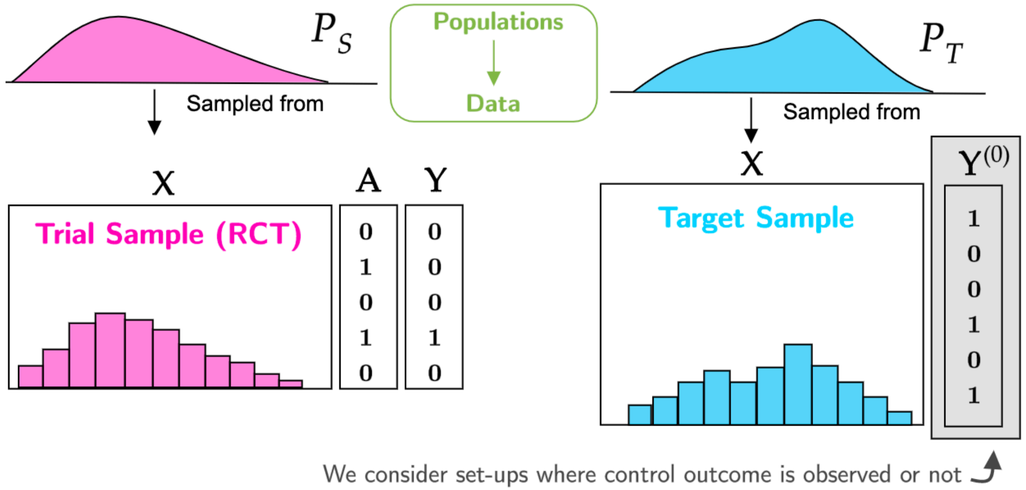




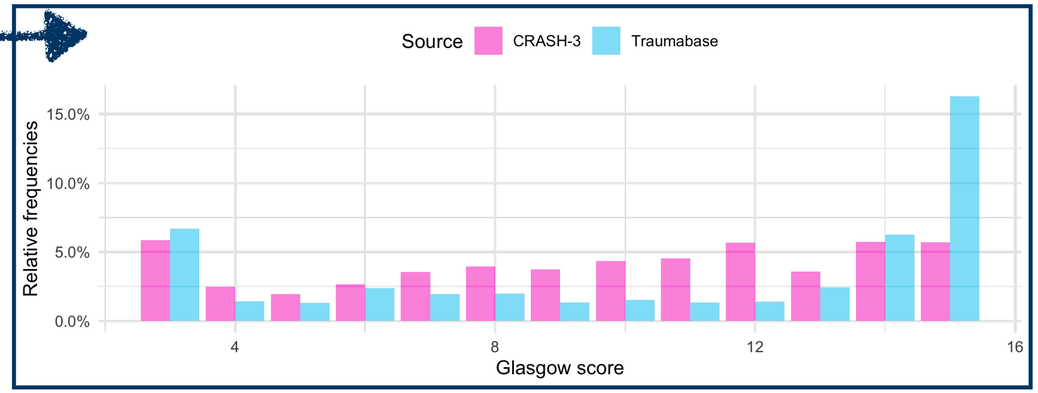
What would be the effect if individuals where sampled in target population?

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

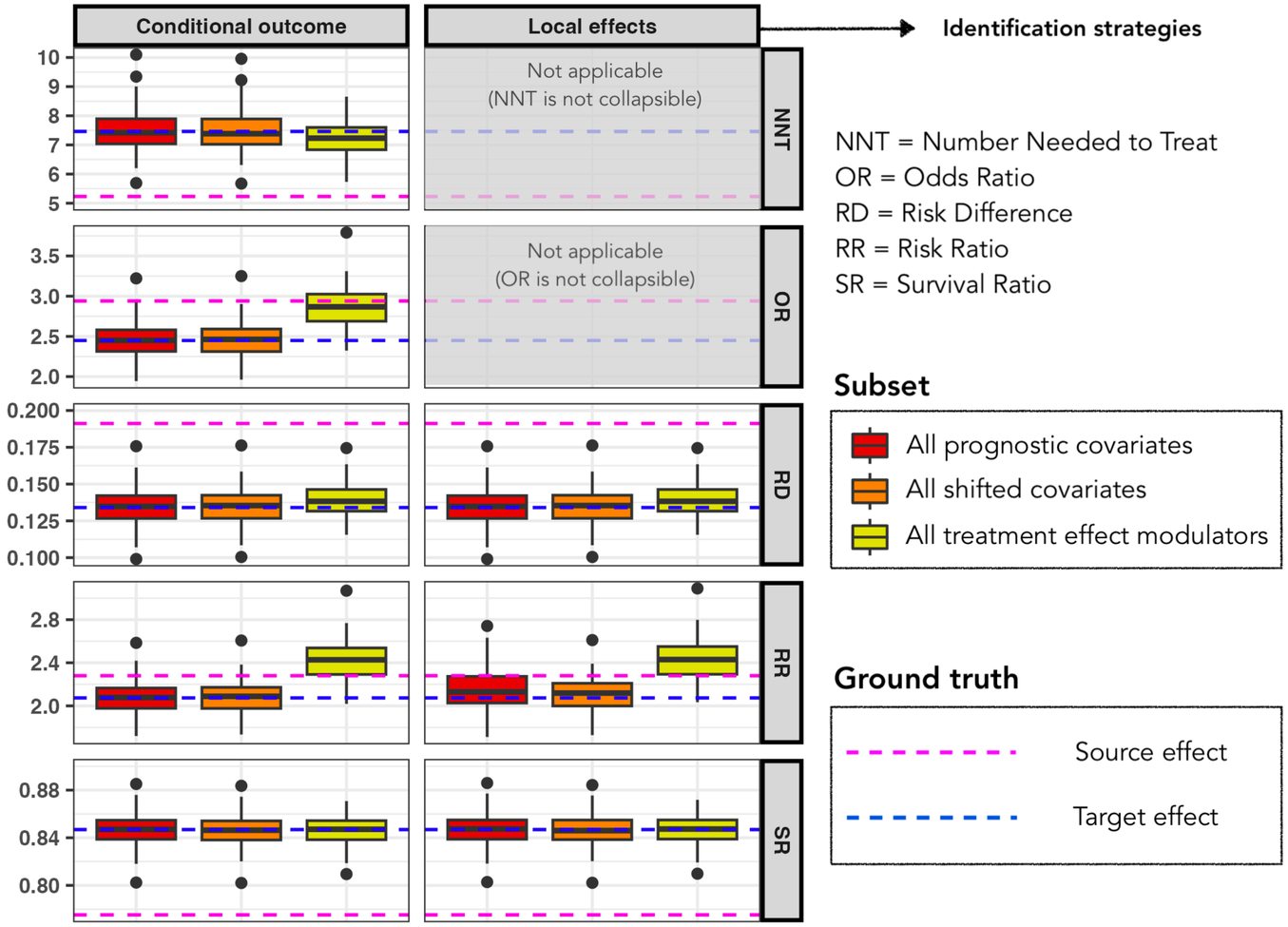
- Probability to die varies
  - Stressed people can die from a heart attack
  - Executioner more merciful when facing women

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

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?

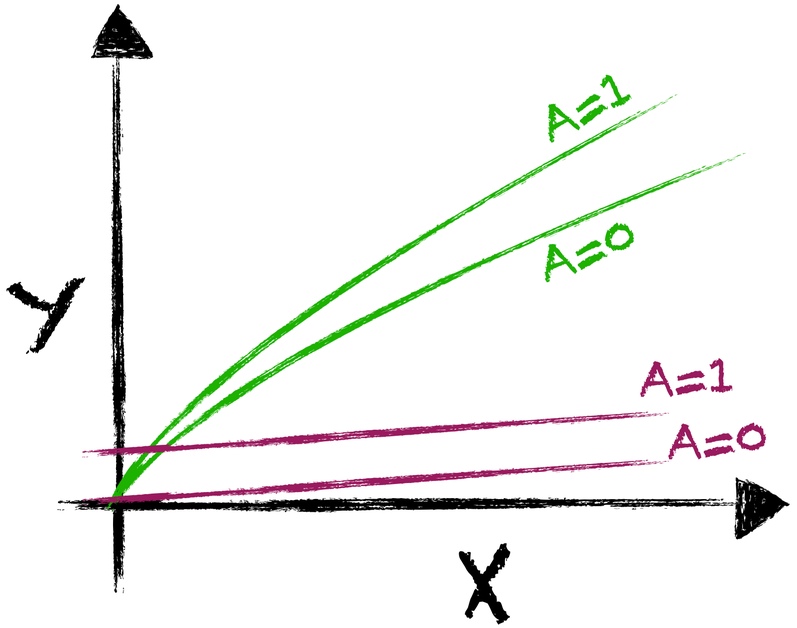
@BenedicteColnet

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

The promise of detailed and larger observational or *real world* data sets

Estimate the efficacy in real-world conditions

- Relying on one data set such as Electronic Health Record or hospital data base

- Emulate a target trial leveraging observed confounding variables

- Solving both representativity and effective treatment given

🎁 *Large sample enabling estimation of stratified effects*

The promise of detailed and larger observational or *real world* data sets

Estimate the efficacy in real-world conditions

- Relying on one data set such as Electronic Health Record or hospital data base

- Emulate a target trial leveraging observed confounding variables

- Solving both representativity and effective treatment given

🎁 *Large sample enabling estimation of stratified effects*

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting

Noa Dagan, M.D., Noam Barda, M.D., Eldad Kepten, Ph.D., Oren Miron, M.A., Shay Perchik, M.A., Mark A. Katz, M.D., Miguel A. Hernán, M.D., Marc Lipsitch, D.Phil., Ben Reis, Ph.D., and Ran D. Balicer, M.D.

ABSTRACT

BACKGROUND

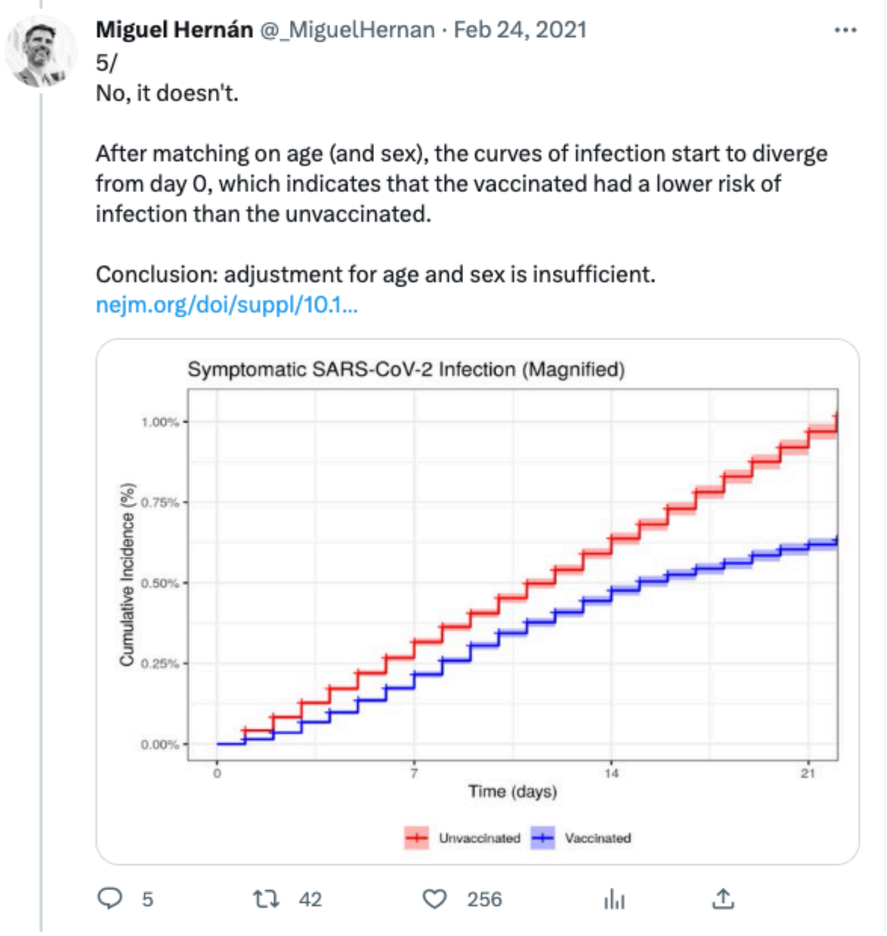
As mass vaccination campaigns against coronavirus disease 2019 (Covid-19) commence worldwide, vaccine effectiveness needs to be assessed for a range of outcomes across diverse populations in a noncontrolled setting. In this study, data from Israel's largest health care organization were used to evaluate the effectiveness of the BNT162b2 mRNA vaccine.

METHODS

All persons who were newly vaccinated during the period from December 20, 2020, to February 1, 2021, were matched to unvaccinated controls in a 1:1 ratio according to demographic and clinical characteristics. Study outcomes included documented infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), symptomatic Covid-19, Covid-19–related hospitalization, severe illness, and death.

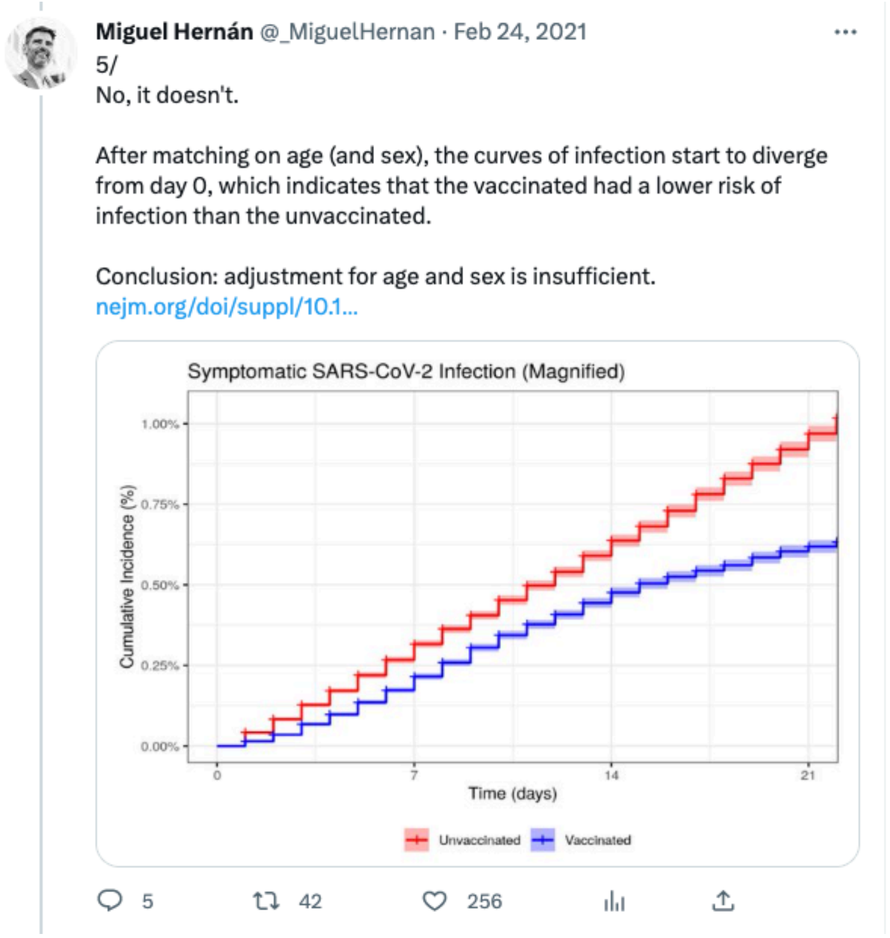
The limits of detailed and larger observational or *real world* data sets

Fear of unobserved confounding



The limits of detailed and larger observational or *real world* data sets

Fear of unobserved confounding



Idea — Using both data sets!

1. Using RCT to check all confounders are observed

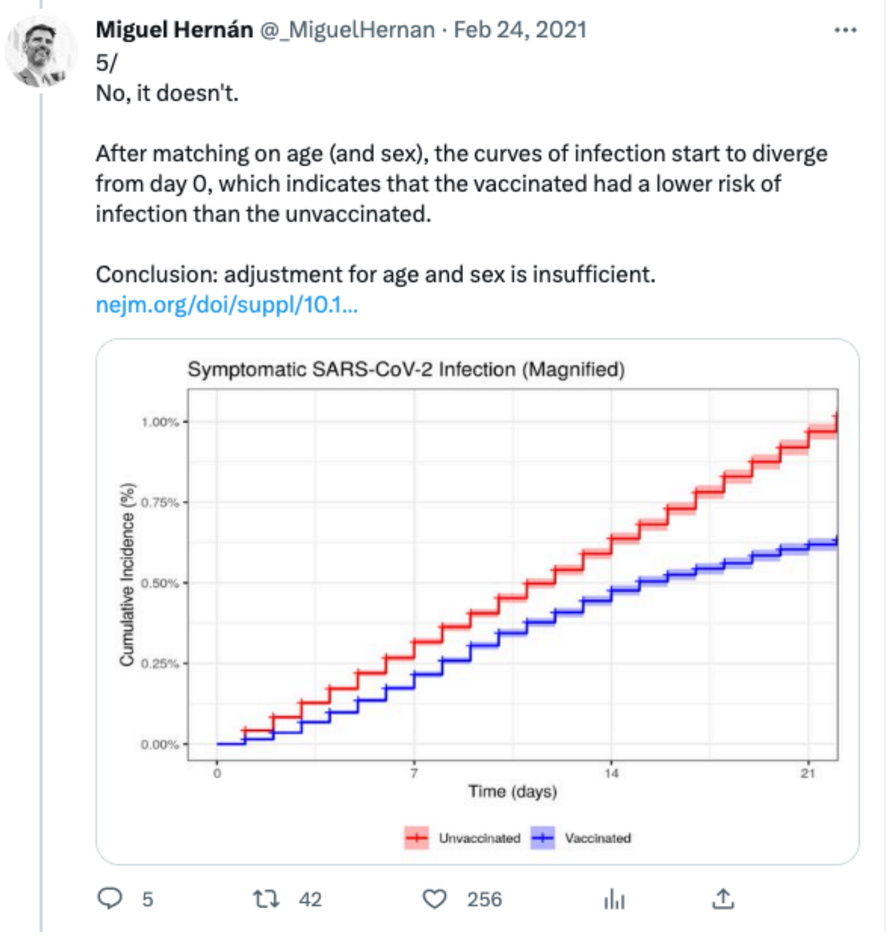
   — **Grounding** observational analysis

2. ***Using observational data to improve trial's representativity***

   — **Generalizing** or **transporting** clinical trial findings toward a new target population

The limits of detailed and larger observational or *real world* data sets

Fear of unobserved confounding



Idea — Using both data sets!

1. Using RCT to check all confounders are observed

   — **Grounding** observational analysis

2. ***Using observational data to improve trial's representativity***

   — **Generalizing** or **transporting** clinical trial findings toward a new target population

Today's focus!