

Challenging binaries in evaluation

Dr Andi Fugard
IFF Research, London, UK

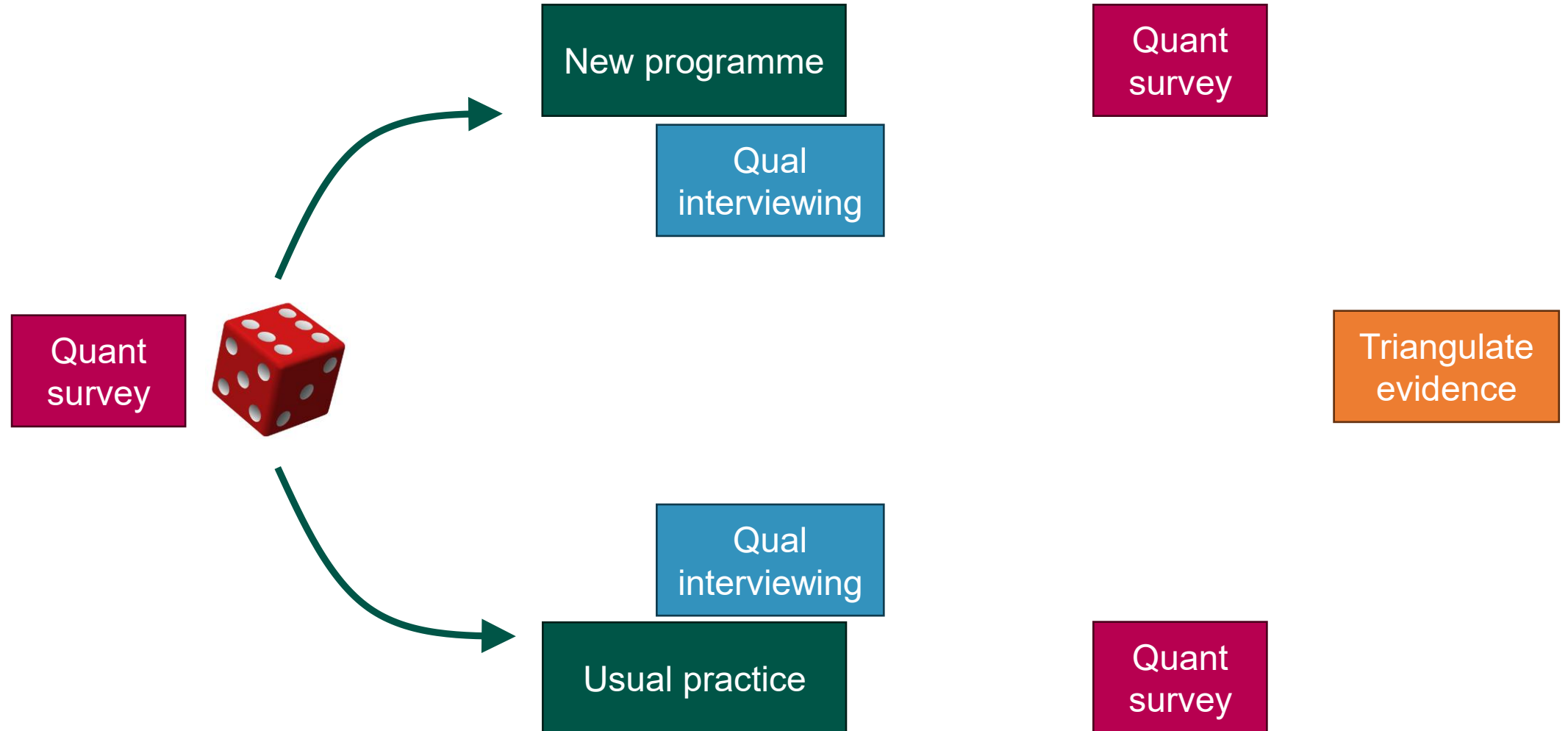
Australian Evaluation Society Conference
Canberra, Ngunnawal and Ngambri Country, Sept 2025

A warmup binary

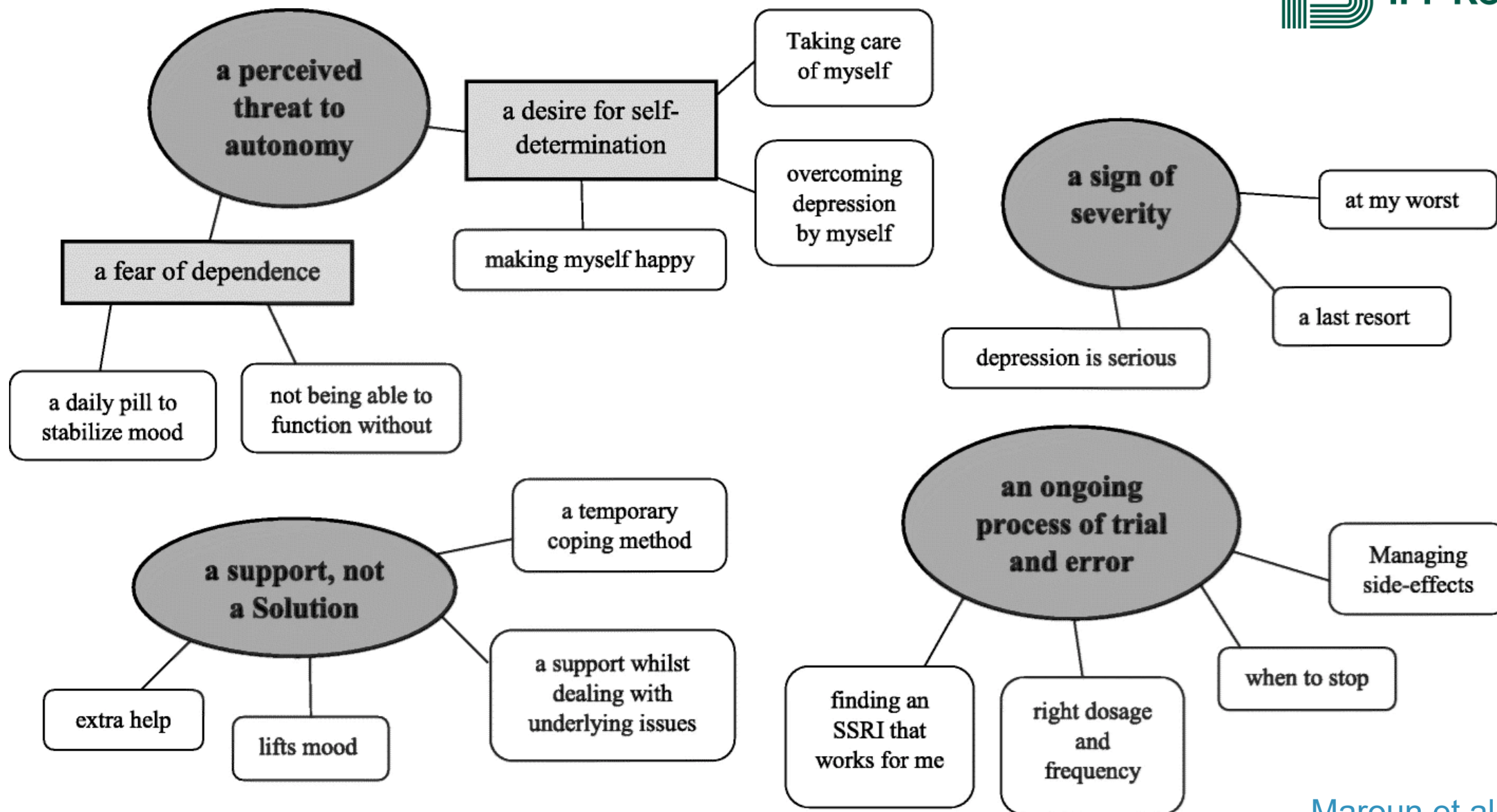
Quantitative

Qualitative

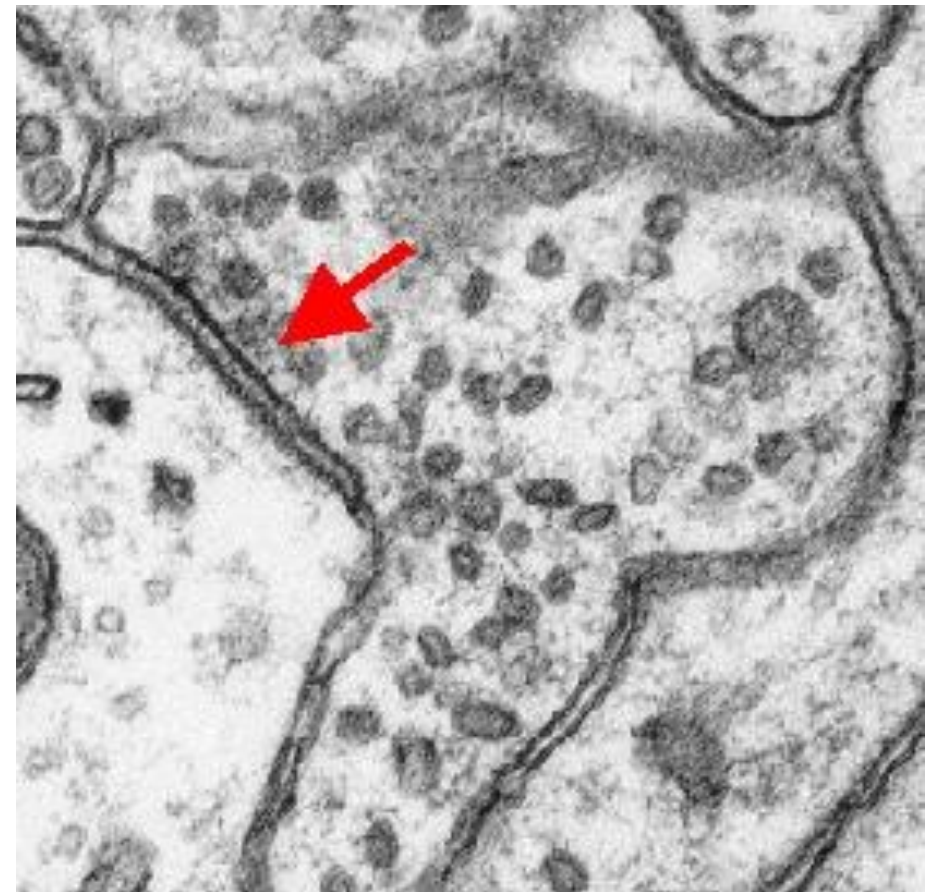
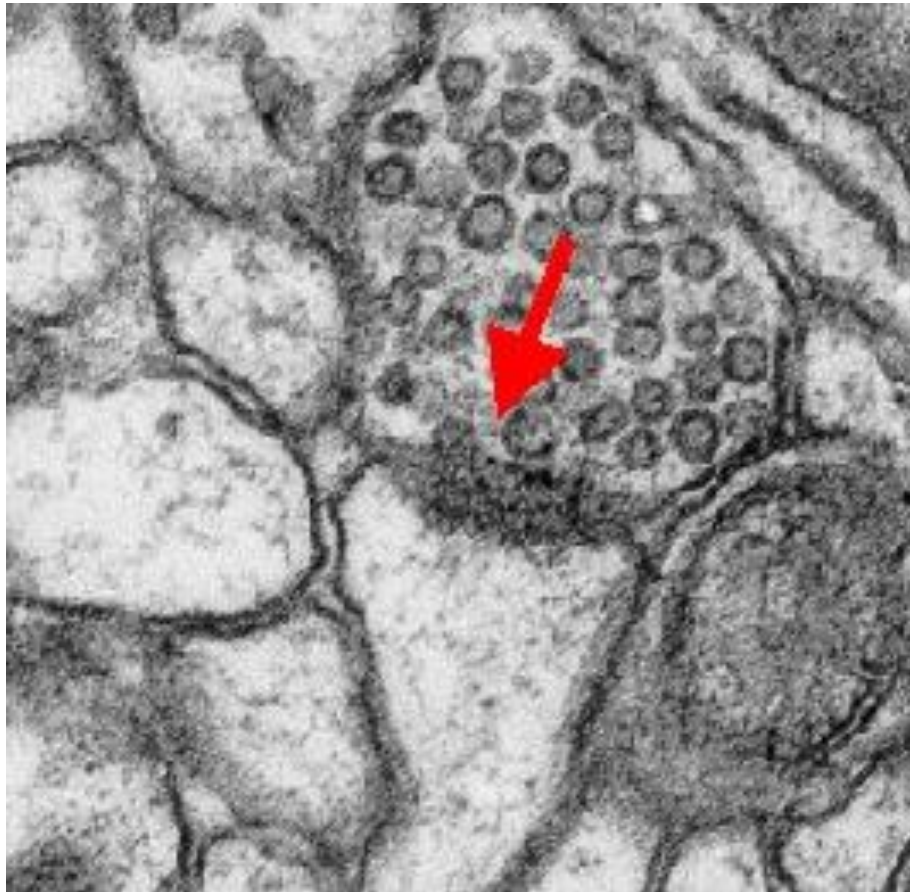
Mixed methods RCTs



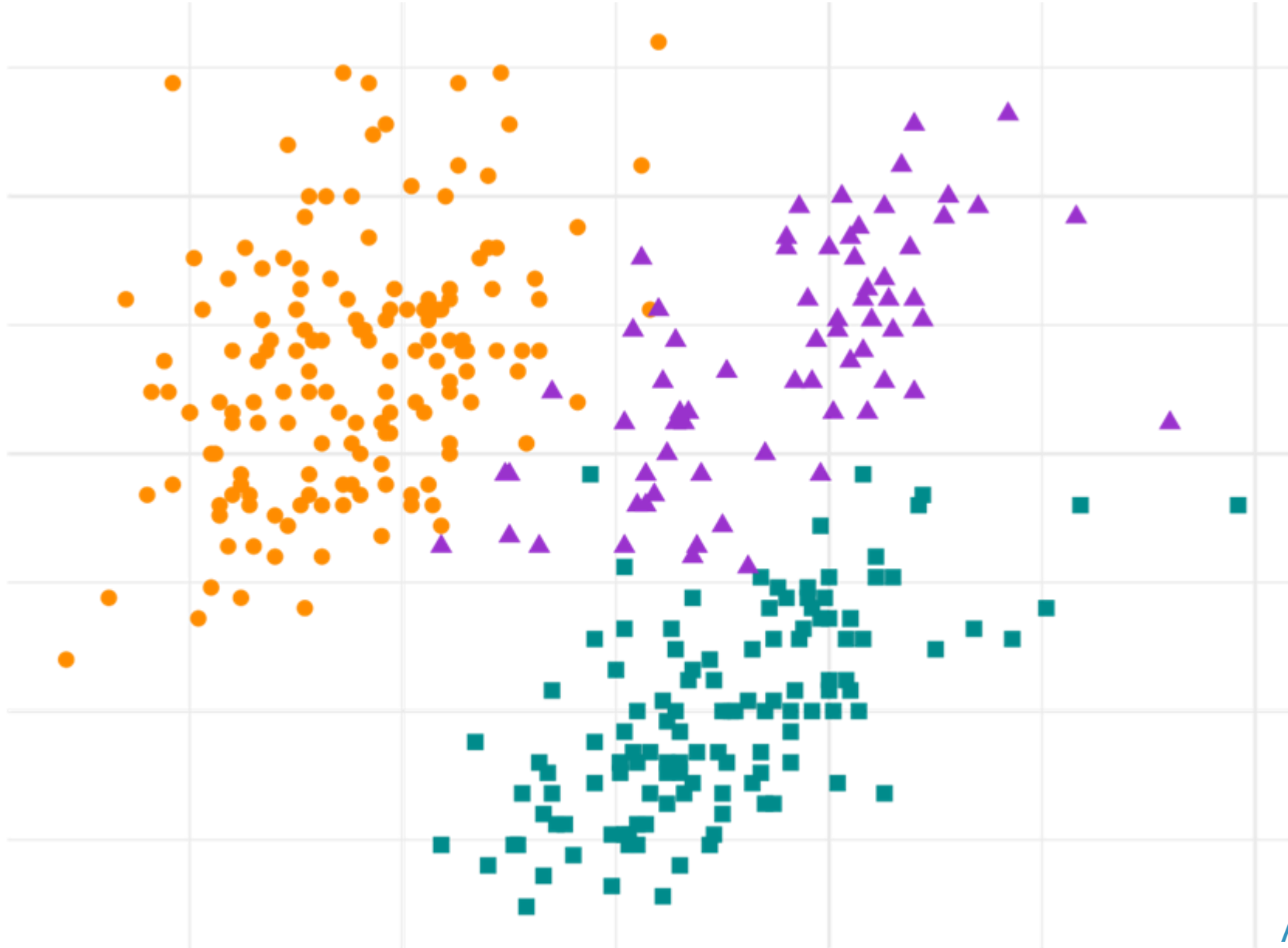
A thematic analysis



Another thematic analysis(?)



You can also do qual analysis of quant data



But we do more with experiences than only classify them

“I used to love it when you said good night to me and we’d try and outdo each other for who had the most quantity of love: love you hundreds, I love you thousands; well I love you millions; and I love yooou trillions. Night, night, don’t let the bed bugs bite; don’t wake the moomims. Night Mum. Mum... And you’d say, ‘all the things we say’. And kiss me and put the night light on and I’d listen for you next door in your and dad’s bedroom. All the things we say.”

Gail Lewis (2009). Birthing Racial Difference: Conversations with my mother and others. *Studies in the Maternal*, 1(1).

Remixed methods

	Purpose	Example approaches
Evoke Empathy	Create (deep, felt) understandings	Storytelling Poetry Theatre
Classify	Identify patterns across material	Thematic analysis Latent class analysis Cluster analysis
Predict	Infer beyond what is currently known	Focus group Statistical modelling Expert group opinion

Another binary

Counterfactual

Non-counterfactual

What is a counterfactual?

Counterfactual conditional

1. If I had got out of bed 10 mins earlier, I wouldn't have missed the train
2. If Jane had been in the comparison group, her maths test result would have been lower

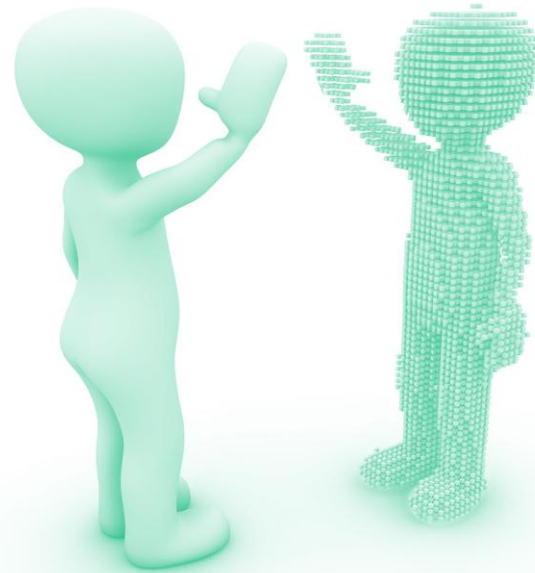
Counterfactual query

1. If I had got out of bed 10 mins earlier, would I have missed the train?
2. If everyone in the intervention group had instead been in the comparison group, what would their average score have been?

Individual causal effects (e.g., Neyman–Rubin)

Actual outcome

Actual outcome
following the
introduction of a
programme

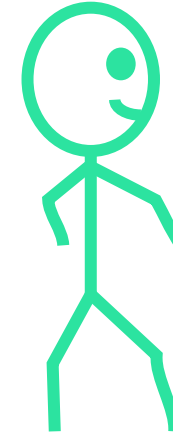
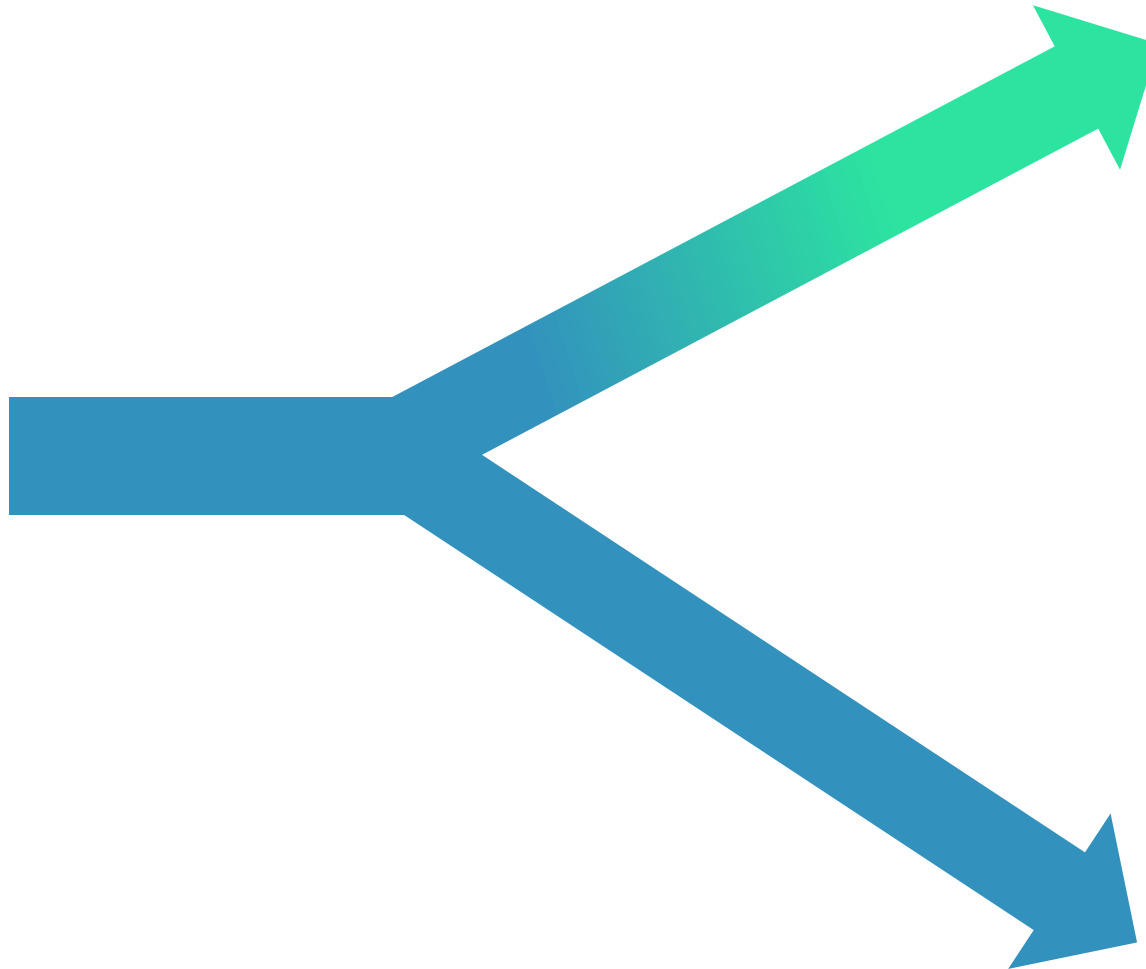
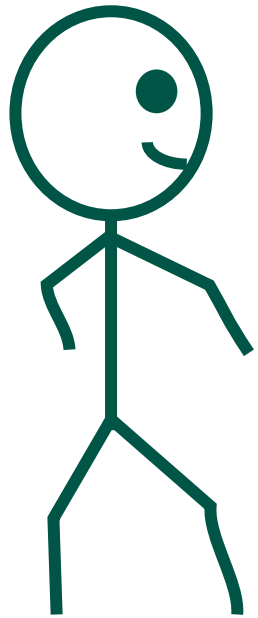


Counterfactual outcome

What the outcome would
have been if, e.g., the
programme had not been
introduced

An **individual's causal effect** is the difference

The “fundamental problem” of causal inference

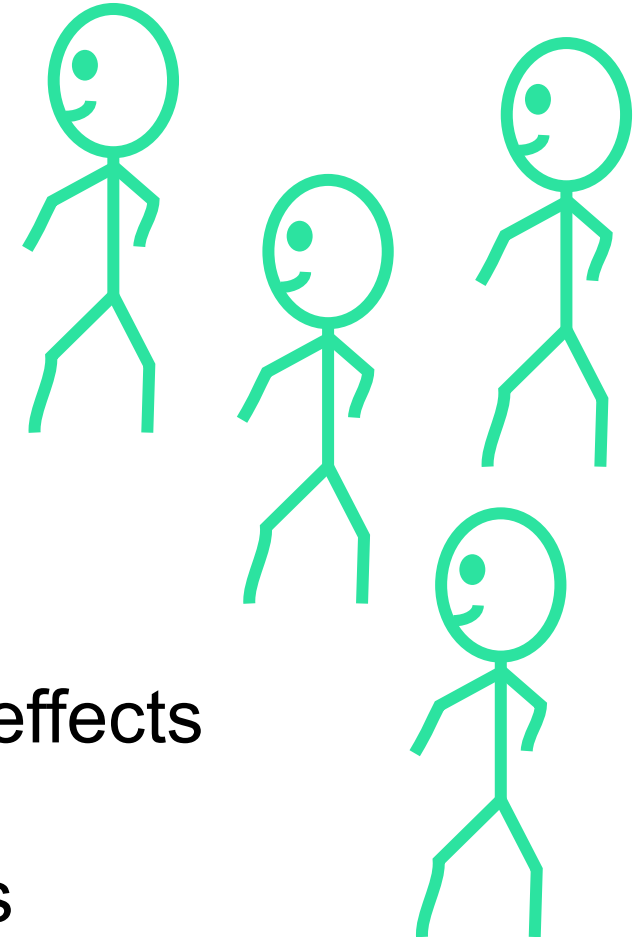
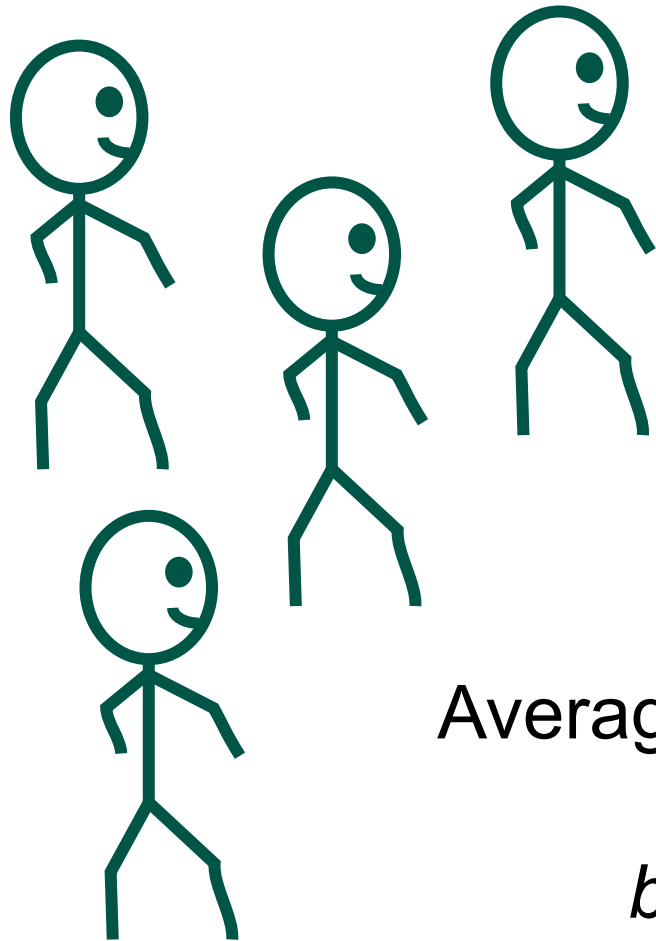


factual



counter-
factual

Counterfactual outcomes can be estimated statistically



Average of *within*-person causal effects
is estimated using
between-person differences

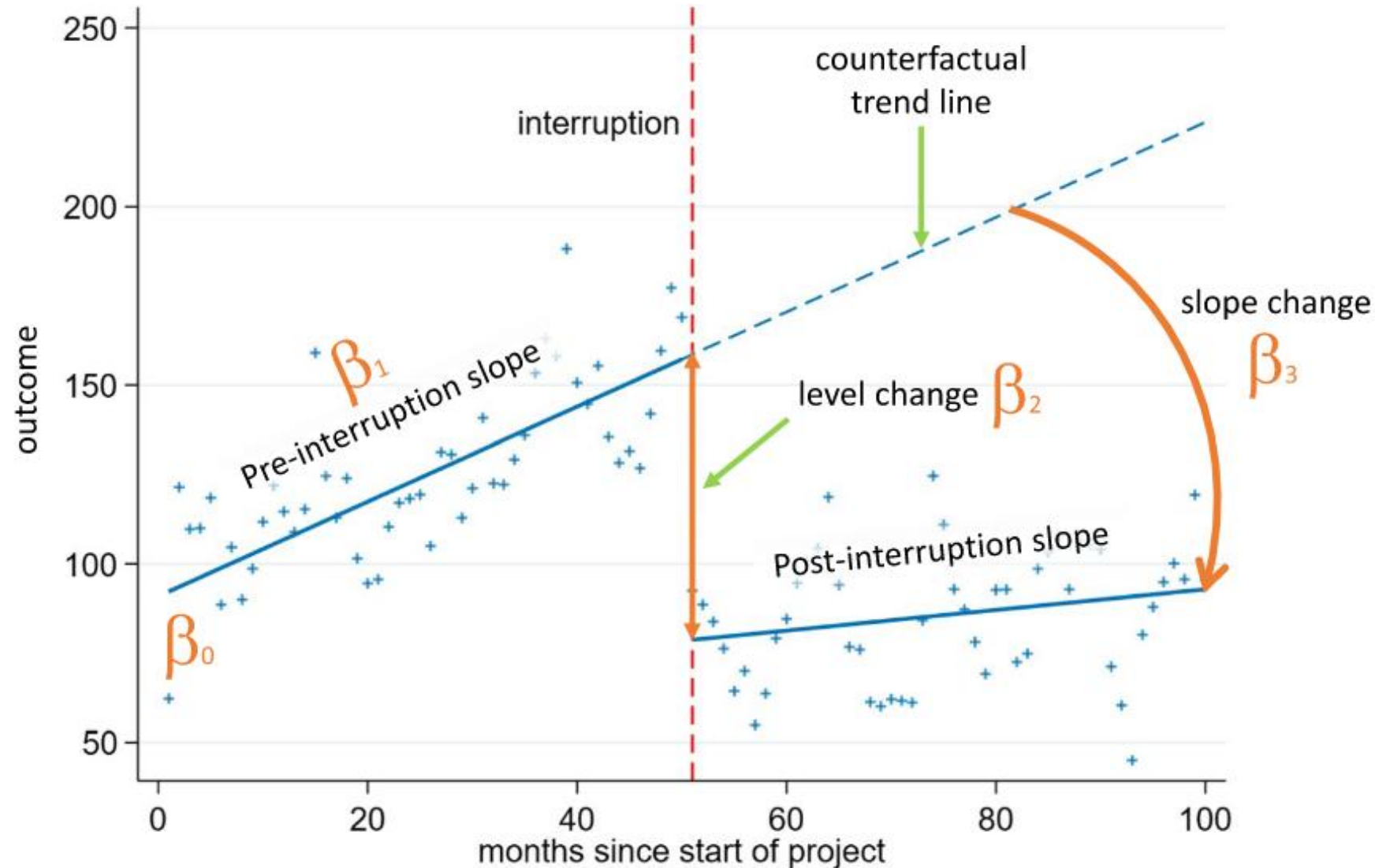
Example assumptions

Sufficiently well-defined intervention and control: we need a precise definition of what we're contrasting

Relevant outcomes data: we need measures of outcomes following conditions that mirror the definitions

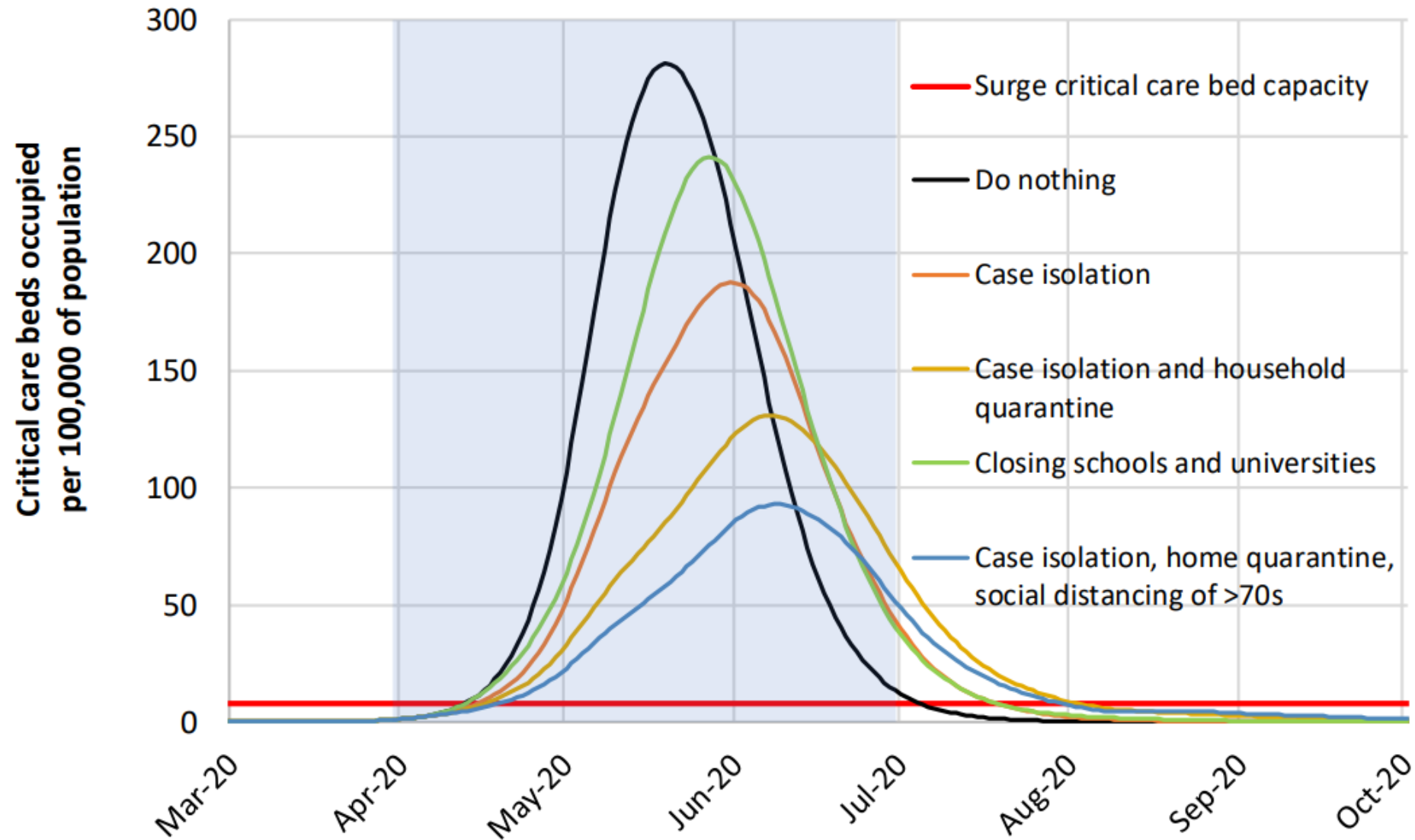
No unmeasured confounding: relevant pre-intervention measures and they are statistically adjusted (or there is random assignment)

Statistical counterfactuals are also possible without a comparison group



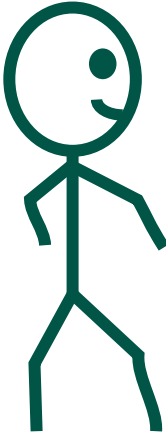
(Turner et al., 2021)

When we know lots about a topic area, we can simulate counterfactuals



Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand

There are also mental models of counterfactuals



If we hadn't received
that grant, the
community centre
would have had to
close

The grant was helpful
and enabled us to
stay open

See the extensive psychology of
reasoning literature
(e.g., Byrne, 2016; Gerstenberg, 2024;
Over & Cruz, 2023)

Evaluation has also been pondering the full breadth of counterfactual evidence for some time; for example:

Counterfactual using before and after analysis

Explicit and latent counterfactuals in narratives

Mueller, Gaus,
& Rech (2014)

Reichardt (2022)

White (2010)

Copestake, Remnant,
& Morsink (2019)

Counterfactual self-
estimation of program
participants

Six non-
RCT/QED
examples

Case study: contribution analysis

“In many situations a counterfactual perspective on causality [...] is unlikely to be useful; experimental designs are often neither feasible nor practical.”

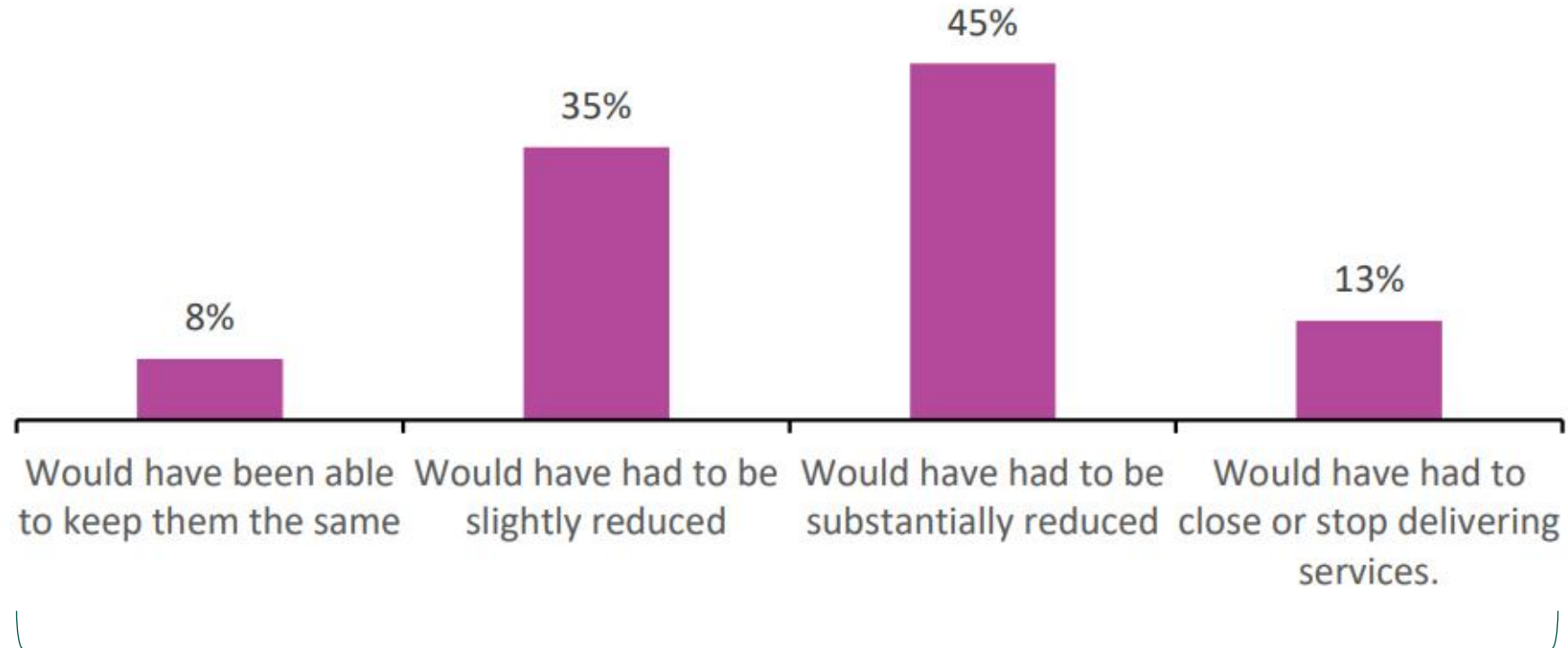
“[Contribution analysis] uses a stepwise (generative) not a counterfactual approach to causality.”

Mayne (2019, pp. 173–4, 176)

Part of a contribution analysis

(Rosa Lau et al., 2022, p. 44)

Figure 4:6 Ability to maintain the same level of service delivery without the VCSE sector funding



A bunch of counterfactuals

Critical appraisal – example considerations

Statistical counterfactual

Check testable assumptions, such as baseline balance

Draw on subject matter expertise to determine whether all relevant data was included in the formal model

Self-reported counterfactual

Check whether narrative explanation is coherent

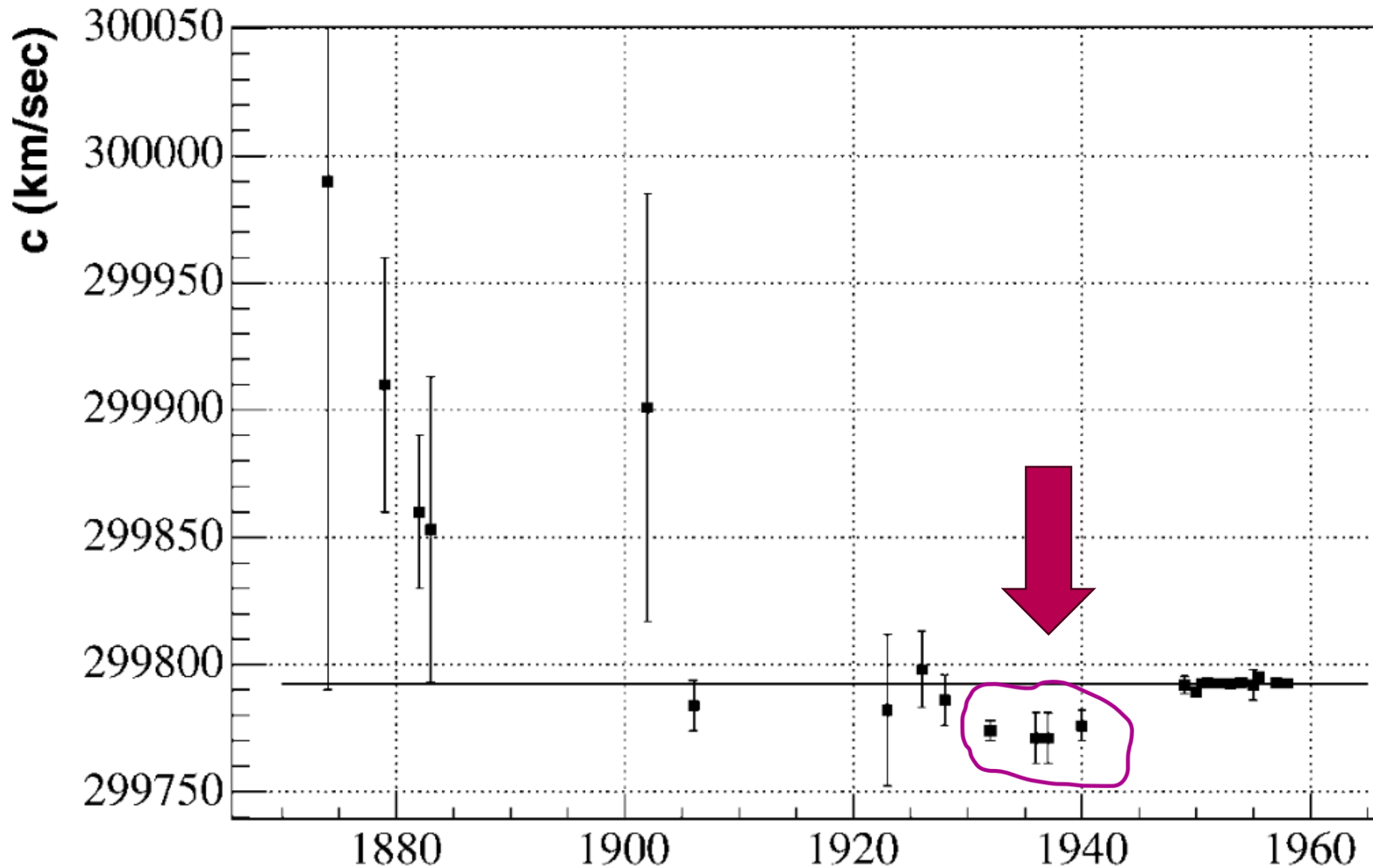
Draw on subject matter expertise to determine whether all relevant data was included in informal model

Final binary for today

Objective

Subjective

The social construction of the speed of light



A famous mathematician on (an approach to) probability

“My thesis, paradoxically, and a little provocatively, but nonetheless genuinely, is simply this:

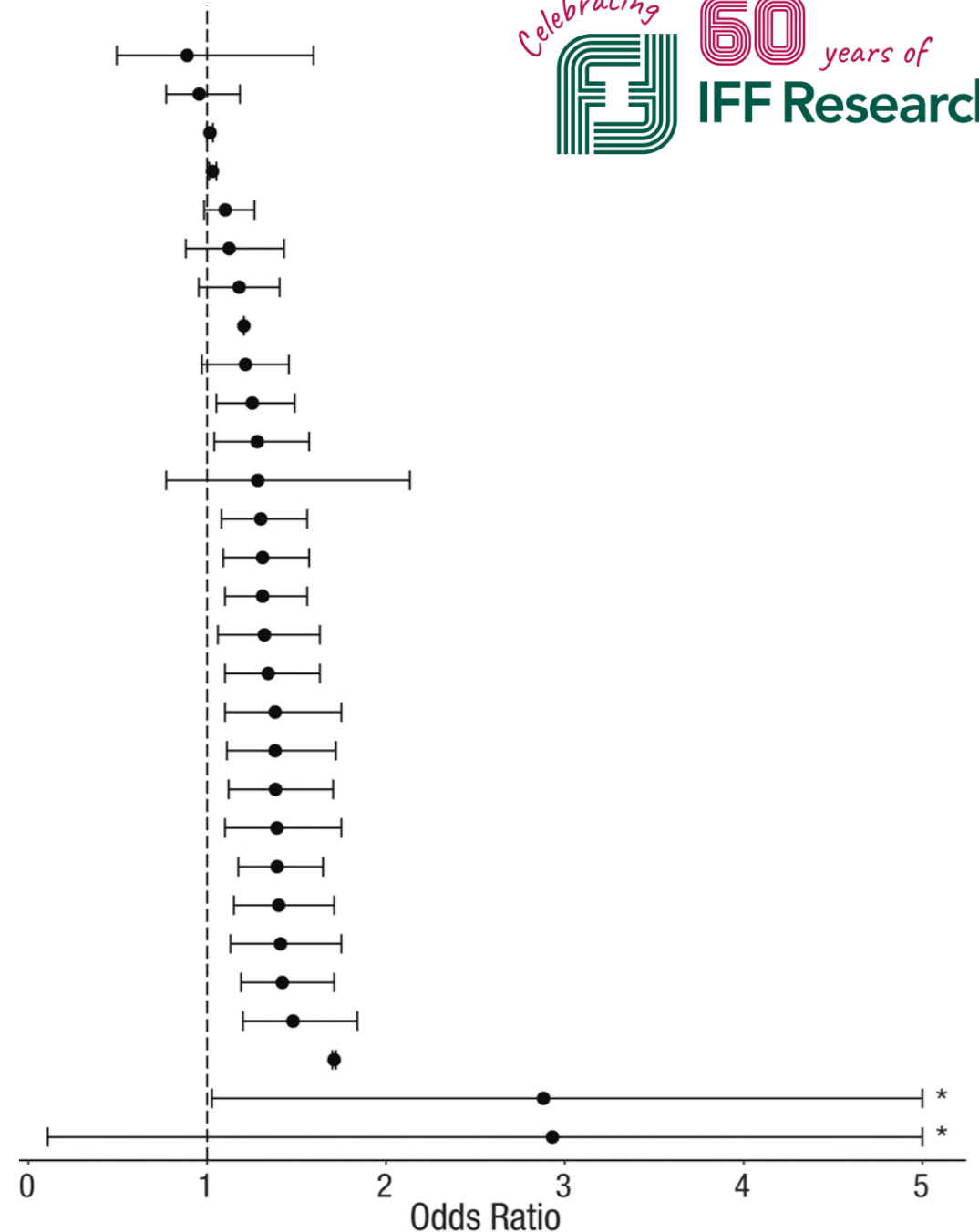
PROBABILITY DOES NOT EXIST

[...] only *subjective* probabilities exist – i.e., the *degree of belief* in the occurrence of an event attributed by a given person at a given instant and with a given set of information.”

Many analysts, one dataset

“Crowdsourcing data analysis, a strategy in which numerous research teams are recruited to simultaneously investigate the same research question, makes transparent how defensible, yet subjective, analytic choices influence research results.”

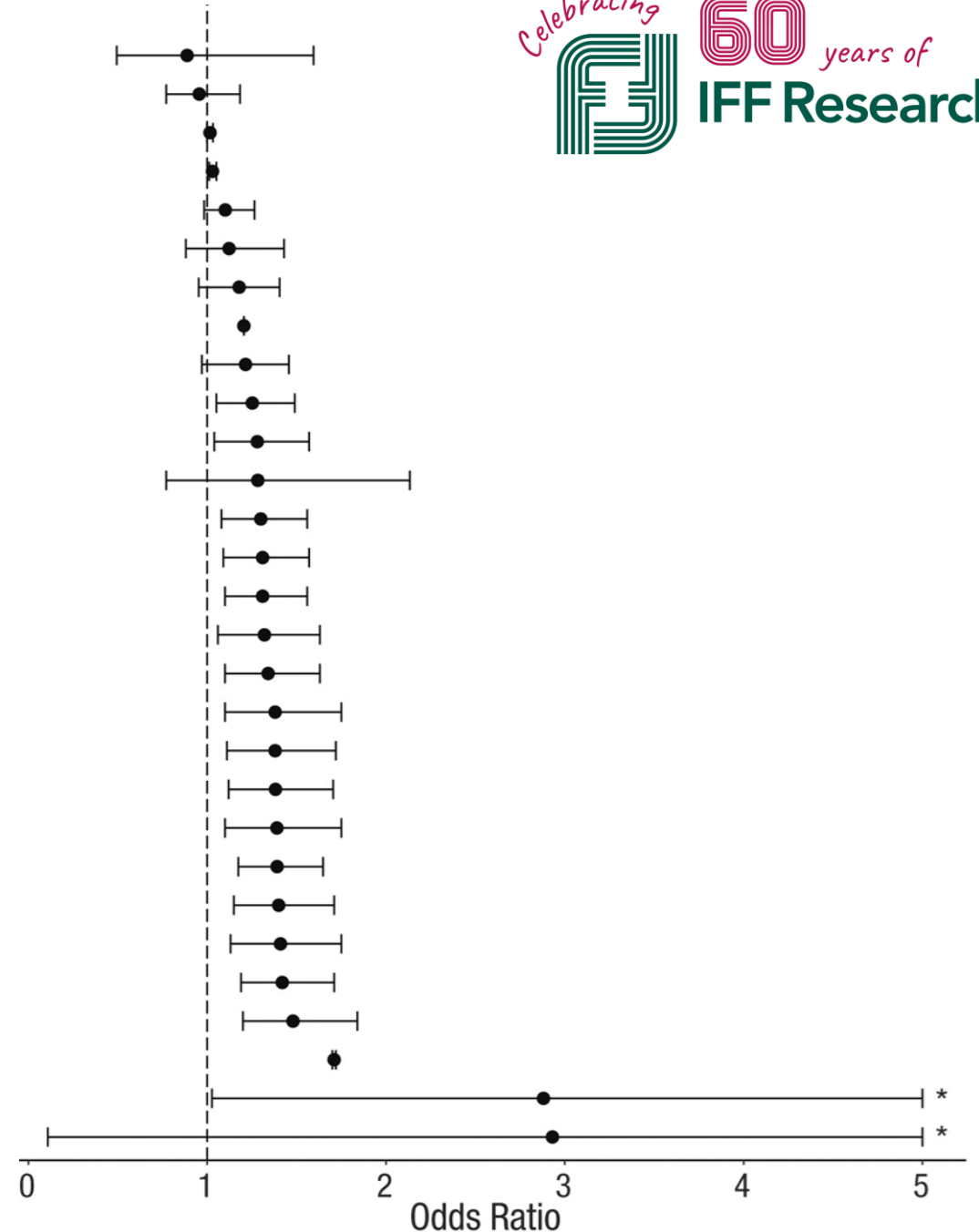
Silberzahn, et al. (2018)



Many analysts, one dataset

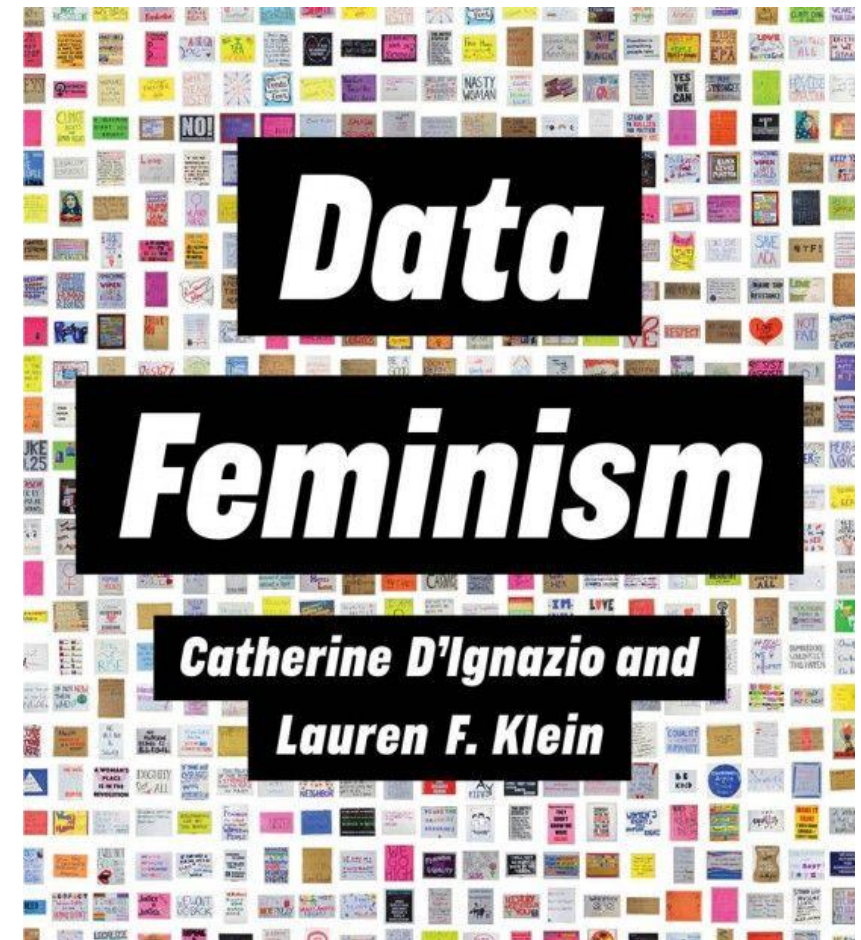
“Crowdsourcing data analysis, a strategy in which numerous research teams are recruited to simultaneously investigate the same research question, makes transparent how defensible, yet subjective, analytic choices influence research results.”

Silberzahn, et al. (2018)



“The problems of gender and racial bias in our information systems are complex, but some of their key causes are plain as day [...]. When data teams are primarily composed of people from dominant groups, those perspectives come to exert outsized influence on the decisions being made—to the exclusion of other identities and perspectives. This is not usually intentional; it comes from the ignorance of being on top. We describe this deficiency as a *privilege hazard*.”

– Catherine D’Ignazio and Lauren F. Klein (2020)





“Raphaël Poulain likes to [...] empty his toolbox, clean it thoroughly, and, finally, to put everything away carefully.”
(Amélie, 2001)



Tel: 020 7250 3035



andi.fugard@iffresearch.com



www.iffresearch.com



IFF Research