Day 2: Policy evaluation and impact assessment

Causal reasoning for policy evaluation

Simon Munzert
Hertie School

The plan for today

Sessions

1. Causal reasoning for policy evaluation
2. Regression and matching
3. Quasi-experiments
4. Spotting crimes against causality

Learning goals

- To practice how to think causal, i.e. practice causal reasoning
- To learn a set of statistical strategies to uncover causal relationships
- To nurture the competence of spotting bad causal evidence

Table of contents

Why causation matters

The causal revolution I: The power of experimentation

Hertie School

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019

© Nobel Media. Photo: A. Mahmoud

Abhijit Banerjee

© Nobel Media. Photo: A. Mahmoud

Esther Duflo

© Nobel Media. Photo: A. Mahmoud

Michael Kremer

"For their experimental approach to alleviating global poverty"

From the press release

"This year's Laureates have introduced a new approach to obtaining reliable answers about the best ways to fight global poverty. In brief, it involves dividing this issue into smaller, more manageable, questions – for example, the most effective interventions for improving educational outcomes or child health. They have shown that these smaller, more precise, **questions are often best answered via carefully designed experiments among the people who are most affected.**

In the mid-1990s, Michael Kremer and his colleagues demonstrated how powerful this approach can be, using field experiments to test a range of interventions that could improve school results in western Kenya. (...) Their experimental research methods now entirely dominate development economics."

The causal revolution II: Causality with observational data

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

© Nobel Prize Outreach. Photo: Paul Kennedy
David Card

© Nobel Prize Outreach. Photo: Risdon Photography
Joshua D. Angrist

© Nobel Prize Outreach. Photo: Paul Kennedy
Guido W. Imbens

"For (...) empirical contributions to labour economics [and] for their methodological contributions to the analysis of causal relationships"

From the press release

"Many of the big questions in the social sciences deal with cause and effect. How does immigration affect pay and employment levels? How does a longer education affect someone's future income? These questions are difficult to answer because we have nothing to use as a comparison. We do not know what would have happened if there had been less immigration or if that person had not continued studying.

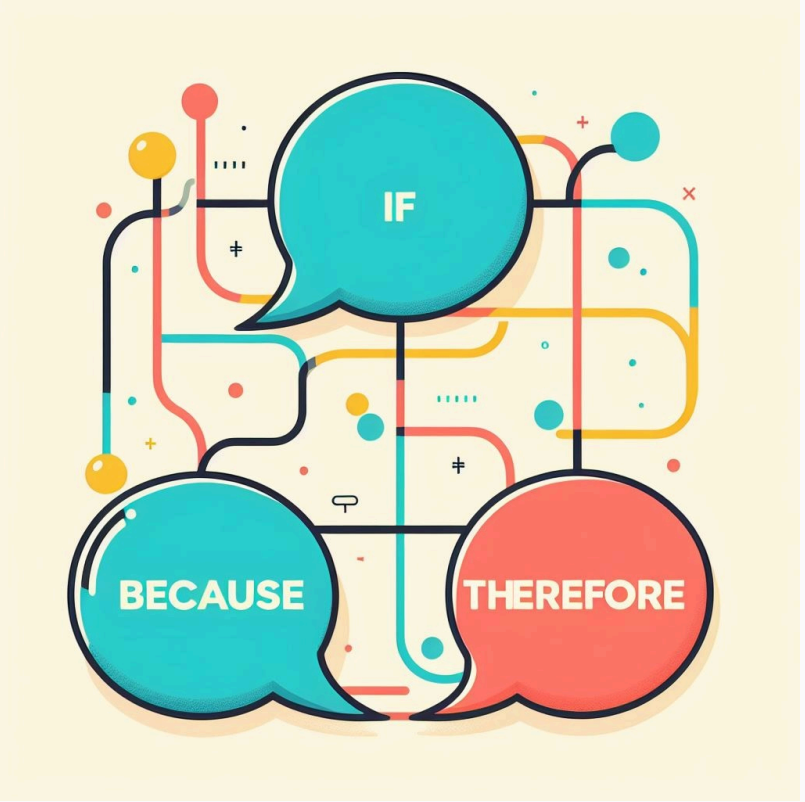
However, this year's Laureates have shown that it is possible to answer these and similar questions using **natural experiments**. The key is to use situations in which chance events or policy changes result in groups of people being treated differently, in a way that resembles clinical trials in medicine."

The language of causal reasoning

Causal vocabulary is all around us

- **What if** everyone (or just a few) had access to universal health care?
- Do social media platforms **affect** election outcomes?
- Can hate speech regulation have unintended **consequences** for people's behavior?
- Will global warming slow down **because of** global measures to reduce C02 emissions?

In the context of impact evaluations, causal language should be used mindfully. In many cases, causal statements are not supported by the design of the study or the data at hand.



Causal reasoning in practice

Implications of research design for causal reasoning

- How to design an **RCT** that cleanly identifies the effect of policy measure X on outcome Y?
- In the absence of an experimental design, how to use **observational data**, e.g., survey data, time-series cross-sectional data and and thus lots of information (variables) on the observations to infer causal effects?
- Should you **control** for covariate Z, or could that actually induce problems?
- What to do when you think you should control for U but **can't observe** it?
- You think X affects Y **through another variable M**. What does that even mean? How could you find out?

All this represents **causal reasoning**. We are going to learn how to apply it systematically.

Counterfactual thinking

Getting started with causation

Causation - a working definition

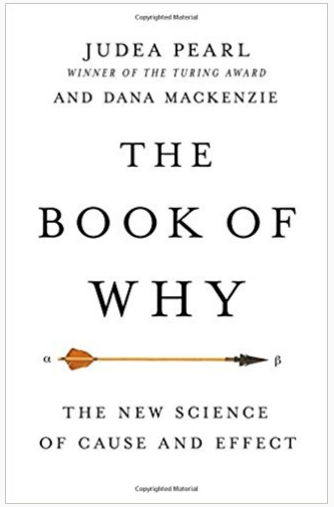
- A.k.a. **causality, cause and effect**
- The idea that one variable [event, process, object; the **cause**] causes another variable [event, process, object, outcome; the **effect**]

Causal reasoning

- The process of studying and identifying causality
- Causal reasoning involves...
  - Ruling out non-causal sources of association and
  - Inferring from the conditions of the occurrence (or absence) of an effect

Causal inference

- In Scott Cunningham's words: "Causal inference is the **leveraging of theory** and deep knowledge of institutional details to **estimate the impact of events and choices on a given outcome of interest**."



Does going to the opera make you live longer?

Researchers in London who followed thousands of people 50 and older over a 14-year period discovered that those who went to a museum or attended a concert just once or twice a year were 14 percent less likely to die during that period than those who didn't.

The chances of living longer only went up the more frequently people engaged with the arts, according to the study, which was published this month in The BMJ, formerly The British Medical Journal. People who went to a museum or the theater once a month or even every few months had a 31 percent reduced risk of dying in that period, according to the study.

The study controlled for socioeconomic factors like a participant's income, education level and mobility, said Andrew Steptoe, a co-author of the study and the head of University College London's research department of behavioral science and health.

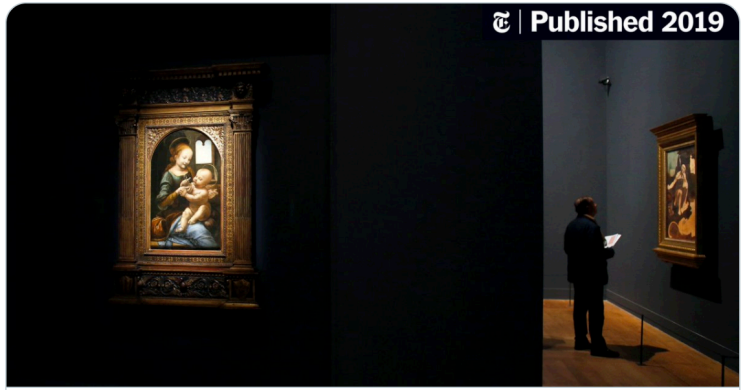
`Source` Cramer, NYTimes (2019)

Does going to the opera make you live longer?

Hertie School

NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



Another Benefit to Going to Museums? You May Live Longer (Published 2019)
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

3:19 PM · Dec 22, 2019 · SocialFlow

318 Retweets **1,705** Quote Tweets **1,239** Likes

RESEARCH

OPEN ACCESS

Check for updates

The art of life and death: 14 year follow-up analyses of associations between arts engagement and mortality in the English Longitudinal Study of Ageing

Daisy Fancourt,[1] Andrew Steptoe[1]

[1]Department of Behavioural Science and Health, University College London, London WC1E 7HB, UK

Correspondence to: D Fancourt
d.fancourt@ucl.ac.uk
(or @Daisy_Fancourt on Twitter;
ORCID 0000-0002-6952-334X)

Cite this as: *BMJ* 2019;367:l6377
http://dx.doi.org/10.1136/bmj.l6377

Accepted: 24 September 2019

ABSTRACT

OBJECTIVE
To explore associations between different frequencies of arts engagement and mortality over a 14 year follow-up period.

DESIGN
Prospective cohort study.

PARTICIPANTS
English Longitudinal Study of Ageing cohort of 6710 community dwelling adults aged 50 years and older (53.6% women, average age 65.9 years, standard deviation 9.4) who provided baseline data in 2004-05.

INTERVENTION
Self reported receptive arts engagement (going to museums, art galleries, exhibitions, the theatre, concerts, or the opera).

MEASUREMENT
Mortality measured through data linkage to the National Health Service central register.

RESULTS
People who engaged with receptive arts activities on an infrequent basis (once or twice a year) had a 14% lower risk of dying at any point during the follow-up (809/3042 deaths, hazard ratio 0.86, 95% confidence interval 0.77 to 0.96) compared with those who never engaged (837/1762 deaths). People who engaged with receptive arts activities on a frequent basis (every few months or more) had a 31% lower risk of dying (355/1906 deaths, 0.69, 0.59 to 0.80), independent of demographic, socioeconomic, health related, behavioural, and social factors. Results were robust to a range of sensitivity analyses with no evidence of moderation by sex, socioeconomic status, or social factors. This study was observational and so causality cannot be assumed.

CONCLUSIONS
Receptive arts engagement could have a protective association with longevity in older adults. This association might be partly explained by differences in cognition, mental health, and physical activity among those who do and do not engage in the arts, but remains even when the model is adjusted for these factors.

Introduction

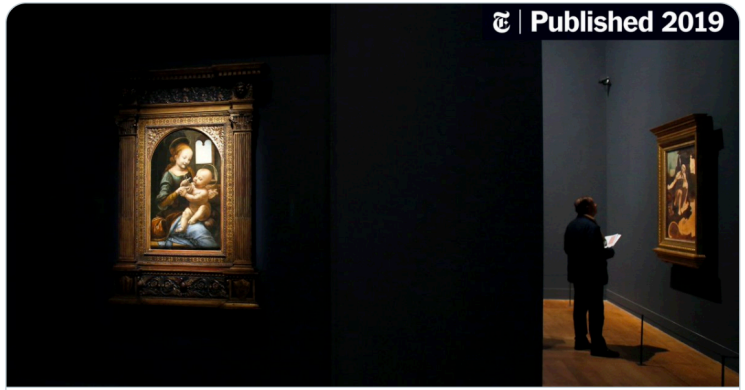
Interest in the salutogenic (health promoting) benefits of the arts is increasing. Arts activities are classified as "multimodal" health interventions because they combine multiple psychological, physical, social, and behavioural factors with an intrinsic aesthetic motivation to engage.[1] While previous studies have shown the association between arts engagement and the prevention and treatment of mental and physical health conditions, including depression, dementia, chronic pain, and frailty,[2-4] whether arts engagement actually confers survival benefits remains unclear. Some research has proposed that the universality of art and the strong emotional responses it induces are indications of its association with evolutionary

`Source` Fancourt, Steptoe / BMJ (2019)

Does going to the opera make you live longer?

NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



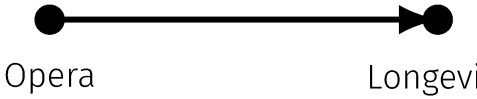
Another Benefit to Going to Museums? You May Live Longer (Published 2019)
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

3:19 PM · Dec 22, 2019 · SocialFlow

318 Retweets **1,705** Quote Tweets **1,239** Likes

Different sources of association between X and Y

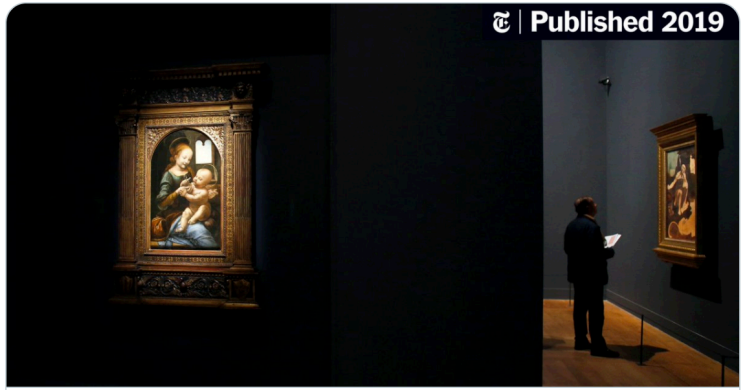
1. X may have caused Y.



Does going to the opera make you live longer?

NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



Another Benefit to Going to Museums? You May Live Longer (Published 2019)
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

3:19 PM · Dec 22, 2019 · SocialFlow

318 Retweets **1,705** Quote Tweets **1,239** Likes

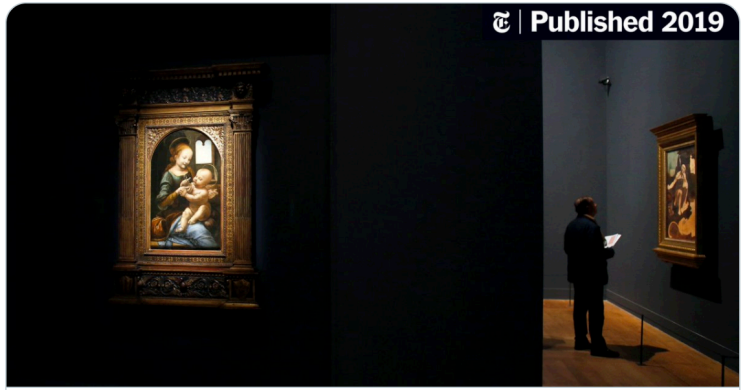
Different sources of association between X and Y

1. X may have caused Y.
2. Y may have caused X.

Does going to the opera make you live longer?

NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



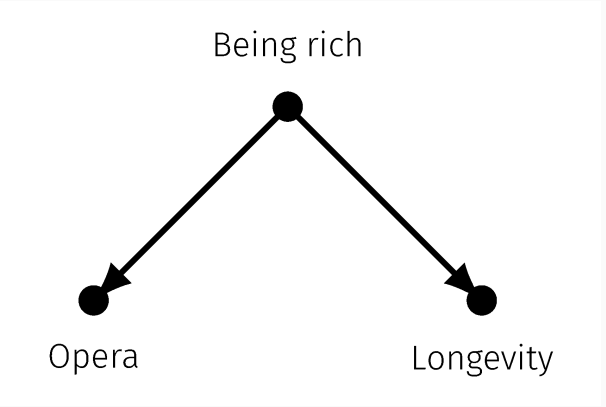
Another Benefit to Going to Museums? You May Live Longer (Published 2019)
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

3:19 PM · Dec 22, 2019 · SocialFlow

318 Retweets **1,705** Quote Tweets **1,239** Likes

Different sources of association between X and Y

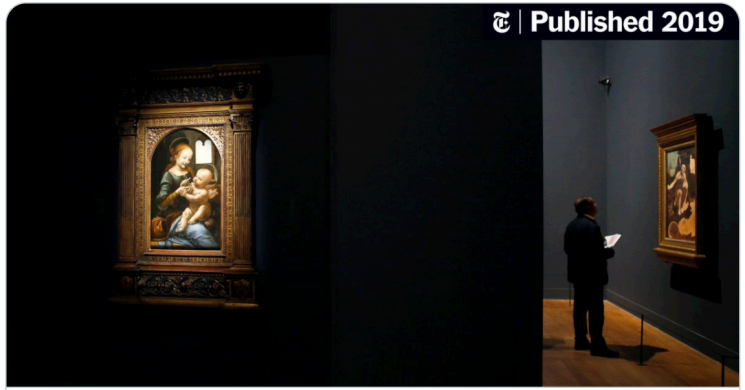
1. X may have caused Y.
2. Y may have caused X.
3. X and Y may share one or more common causes.



Does going to the opera make you live longer?

NYT Health
@NYTHealth

Want to live longer? Try going to the opera. Researchers in Britain have found that people who reported going to a museum or concert even once a year lived longer than those who didn't.



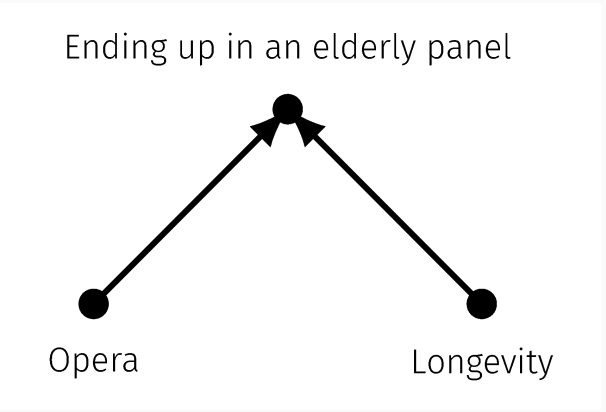
Another Benefit to Going to Museums? You May Live Longer (Published 2019)
Researchers in Britain found that people who go to museums, the theater and the opera were less likely to die in the study period than those who didn't.
nytimes.com

3:19 PM · Dec 22, 2019 · SocialFlow

318 Retweets **1,705** Quote Tweets **1,239** Likes

Different sources of association between X and Y

1. X may have caused Y.
2. Y may have caused X.
3. X and Y may share one or more common causes.
4. The association may have been induced by conditioning on a common effect of X and Y.



Does going to the opera make you live longer?

Hertie School

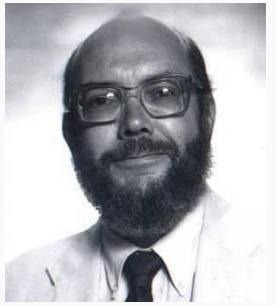
Different sources of association between X and Y

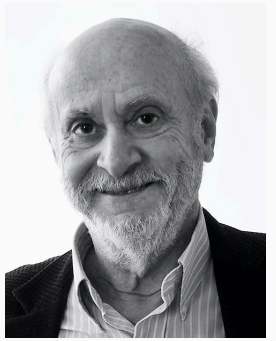
1. X may have caused Y.
2. Y may have caused X.
3. X and Y may share one or more common causes.
4. The association may have been induced by conditioning on a common effect of X and Y.
5. Chance.

Thinking in terms of counterfactuals

The Neyman-Rubin causal model

- The Neyman-Rubin causal model (or just Rubin causal model, RCM) is a formal approach to do causal inference that fueled the potential outcomes framework (POF)
- Up until today, it's the mainstream framework to (statistically) talk about causation





The idea of potential outcomes

- The POF assumes that each subject has a **potential outcome under both treatment states**
- Evidently, each subject can only be observed in one treatment state at any point in time
- Therefore, the individual treatment effect is a non-observable counterfactual ("the fundamental problem of causal inference", Holland 1986)
- The idea is to think about the **counterfactual** outcome that would have been observed had the treatment been different

Thinking in terms of counterfactuals

Potential outcomes under treatment and control

- Assume we have a binary treatment variable X (e.g., going to the opera)
- We observe the outcome Y (e.g., longevity)
- The potential outcomes are $Y(1)$ and $Y(0)$

Group	Longevity (under treatment)	Longevity (under control)
Going to opera	Observable as Y	Counterfactual
Not going to opera	Counterfactual	Observable as Y

The fundamental problem of causal inference

- Why is the causal effect of X on Y not simply the difference between the upper-left and lower-right cell in the table?
- We can only observe one of the two potential outcomes!
- Causal inference implies answering **'What if'** questions, i.e. imagine a counterfactual world in which the treatment was different for a given unit

Addressing the dilemma

- We can try and design studies in a way so that the expected outcome under a treatment state are the same for all units
- Or, we can try to (statistically) adjust such that the observed data is as close as possible to the counterfactual data

Causal designs: experimental and observational

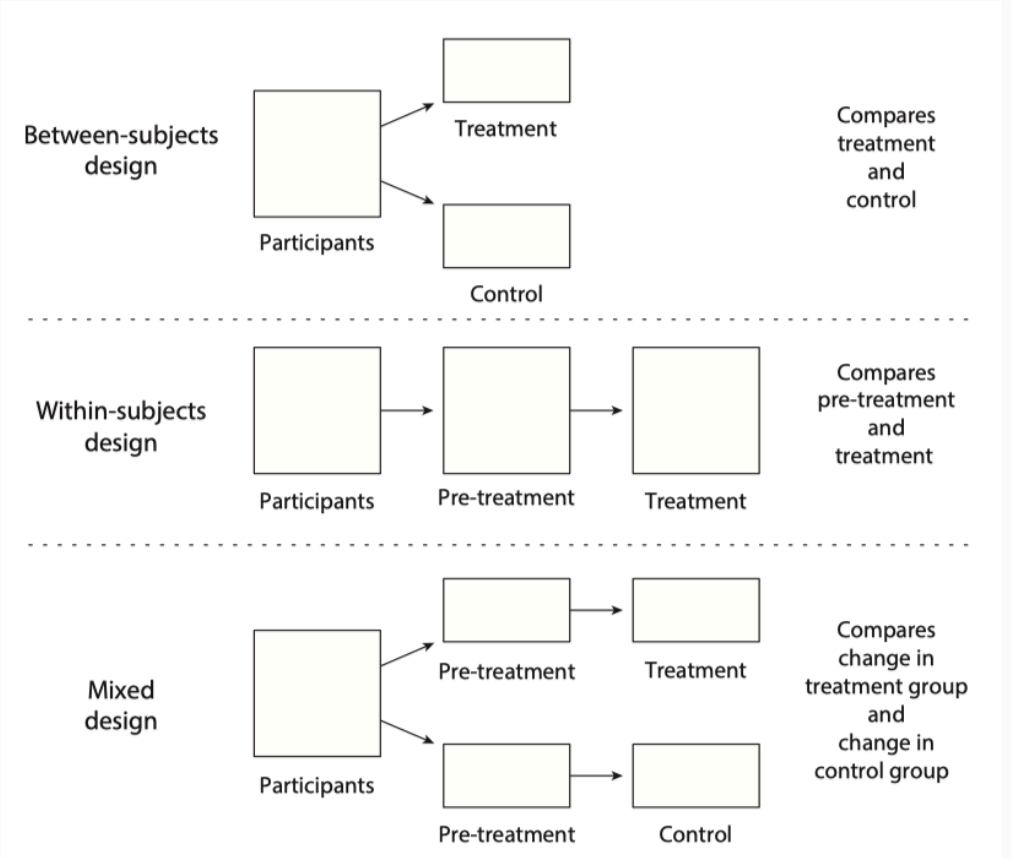
Why experiments?

Dilemmas of observational research

- When does correlation imply causation?
- How can we be sure we did not overlook an important cause?

The promises of experimental designs

- Randomization!
- Researchers set the treatment, but chance decides who gets it and who doesn't
- By randomly assigning the treatment, it is designed to be statistically independent from any other variable.



`Source` Matt Salganik, Bit by Bit

Issues to consider when running (and judging) experiments

1. Balance: Are treatment and control groups balanced on relevant covariates (potential confounders)?

- Run mean comparisons and significance tests
- Condition on pre-treatment covariates
- Ideally, condition on pre-treatment outcomes

2. Power: Is your sample big enough to uncover an effect of a certain size?}

- The higher the statistical power for a given experiment, the lower the probability of making a Type II (false negative) error. Power = 1 - Type II error
- Run a power analysis (ideally before data collection)
- If you can afford it, adapt sample size and/or design on the basis of your power calculations

Issues to consider when running (and judging) experiments

3. Compliance: Do participants participate as intended?

- Often we cannot force subjects to take specific treatments, and **noncompliance** is an issue
- Intent-to-Treat (ITT) effect $\neq$ treatment effect
- Selection bias: self-selection into treatment/control groups

	Drinks no coffee	Drinks coffee	Total
Assigned to Control	80	20	100
Assigned to Treatment	10	90	100
Total	90	110	200

- Encouragement design: Randomize the encouragement to receive the treatment rather than the receipt of the treatment itself

4. External validity: How does the experimental setting connect to the "real world"?

- Is the treatment comparable to interventions you want to compare it to?
- Is the outcome an actual measure of interest?
- How "different" is your sample from the target population on relevant characteristics?

Example: Tracking and promoting usage of a COVID-19 contact tracing app

nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-01044-x

Check for updates

Tracking and promoting the usage of a COVID-19 contact tracing app

Simon Munzert[1]✉, Peter Selb[2], Anita Gohdes[1], Lukas F. Stoetzer[3] and Will Lowe[1]

Digital contact tracing apps have been introduced globally as an instrument to contain the COVID-19 pandemic. Yet, privacy by design impedes both the evaluation of these tools and the deployment of evidence-based interventions to stimulate uptake. We combine an online panel survey with mobile tracking data to measure the actual usage of Germany's official contact tracing app and reveal higher uptake rates among respondents with an increased risk of severe illness, but lower rates among those with a heightened risk of exposure to COVID-19. Using a randomized intervention, we show that informative and motivational video messages have very limited effect on uptake. However, findings from a second intervention suggest that even small monetary incentives can strongly increase uptake and help make digital contact tracing a more effective tool.

As governments around the globe are seeking targeted exit strategies from lockdown measures to contain coronavirus disease 2019 (COVID-19), evidence of widespread pre-symptomatic transmission and short generation times[1,2] calls into question traditional containment measures based on symptomatic surveillance. An early influential modelling study[3] suggested controlling the epidemic with digital contact tracing (DCT); that is, mobile apps that log and report encounters between infected persons and mobile users to prevent onward transmission. Apparently successful rollouts in Singapore and South Korea[4] have, at the time of writing (17 December 2020), encouraged more than 40 countries to introduce DCT apps[5].

Broad uptake in the population is considered key for DCT app effectiveness; an influential study suggested that a 60% adoption rate would be sufficient[3], although the simulations excluded supporting measures such as facial masks and social distancing[6]. Minimizing uptake differentials across social groups is also important. Not only does unequal access to smartphone technology exacerbate existing inequities and raise ethical concerns[7], but the overall effectiveness of DCT apps depends on the users' contact network structures and mixing behaviour[8,9]. For instance, app users might practise data on app usage with a survey capturing individual measures of a rich set of subject characteristics.

Results

Our study was designed to both track and stimulate usage of the official German DCT app over a period of ~100 d (see Supplementary Discussion for background information). Figure 1 illustrates the study setup (see Methods for a more detailed description of subject recruitment and panel design). Our initial survey sample included $n = 2{,}044$ individuals recruited from a commercial access panel using quotas to reflect the age, gender and education distribution of the adult online population resident in Germany (see Supplementary Information). The provider also tracked the online behaviour of a volunteer panel of $n = 1{,}132$ individuals, 649 of whom participated in the survey. Our study population therefore consisted of three subgroups: a tracking-only group for which we had sparse but high-quality behavioural data ($n = 482$) and which we used as a baseline control for the analysis of tracked app usage; a survey-only group for which we had rich self-reports but no behavioural data ($n = 1{,}395$); and a survey-tracking group for which we had both self-reports and behavioural data ($n = 649$).

https://go.nature.com/3auwDjA

Study design

Hertie School

App Tracking Data

Members of the survey provider's passive tracking panel are incentivized to provide mobile app usage histories via passive metering software Wakoopa.

Survey Wave 1

Participants complete a 20-minute survey about sociodemographic, attitudinal and behavioral characteristics.

Message Stimulus

As part of the initial survey, participants are randomly assigned to one of two treatment conditions or the control condition with equal probability.

Survey Wave 2

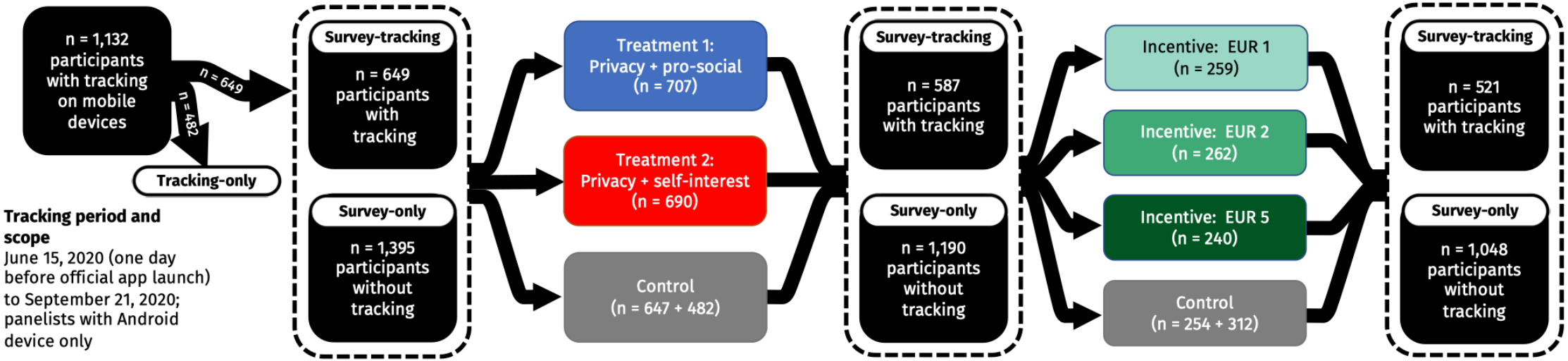
On average 12 days after the initial survey, participants are re-invited to a follow-up survey in which key measures of attitudes and behaviors are repeated.

Incentivization

As part of the follow-up survey, self-reported non-users of the app are randomly assigned to one of three incentivization conditions or the control condition with equal probability.

Survey Wave 3

On average 28 days after the survey Wave 2, participants are re-invited to another follow-up survey in which key measures of attitudes and behaviors are repeated.



Tracking period and scope
June 15, 2020 (one day before official app launch) to September 21, 2020; panelists with Android device only

Data
Corona app usage, time stamps, duration, device information

Tracking baseline
For analyses of tracked app usage, the 482 mobile Tracking-only sample participants are used as additional control units

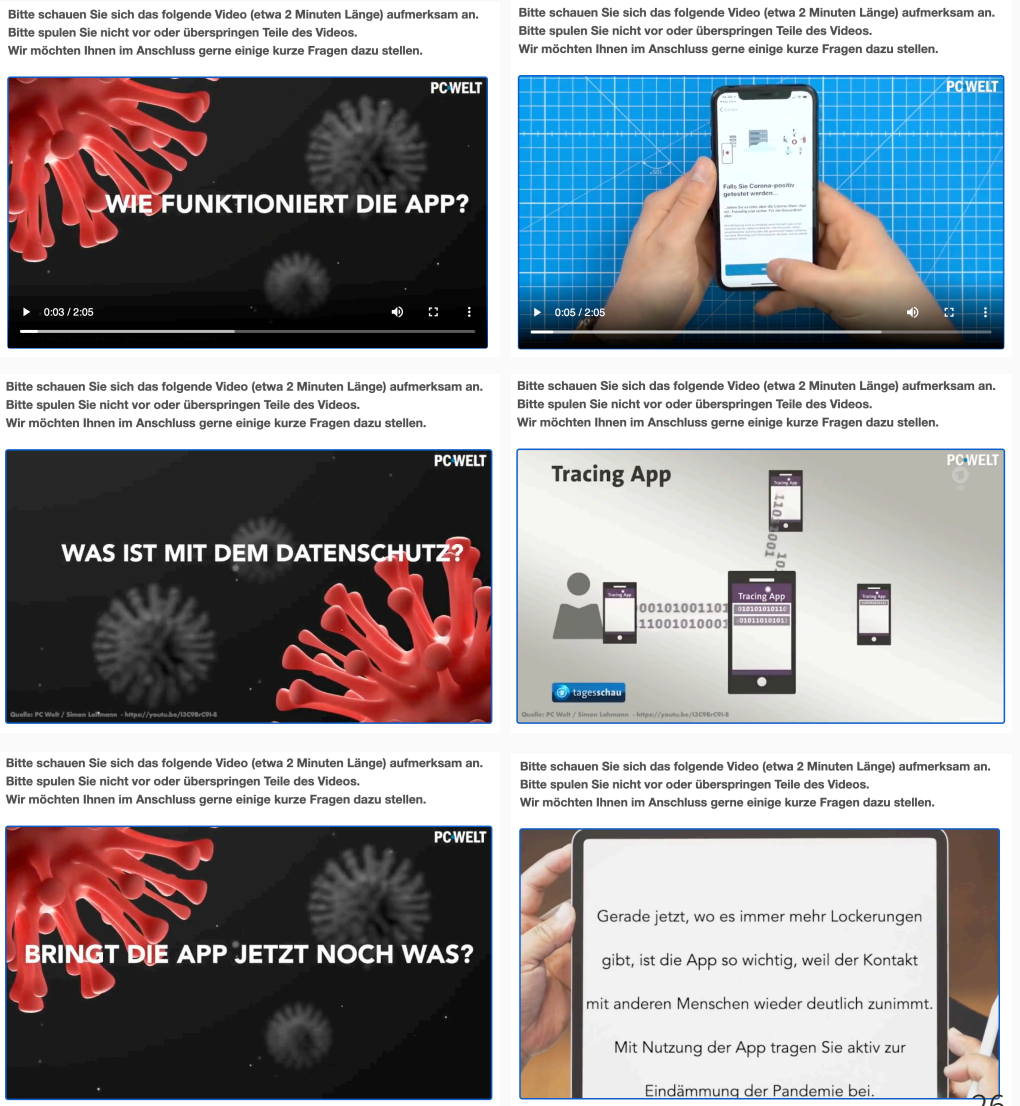
Tracking baseline
For analyses of tracked app usage, the 312 mobile Tracking-only sample participants who did not have the app installed at the time of Follow-Up Survey I are used as additional control units

Experimental intervention

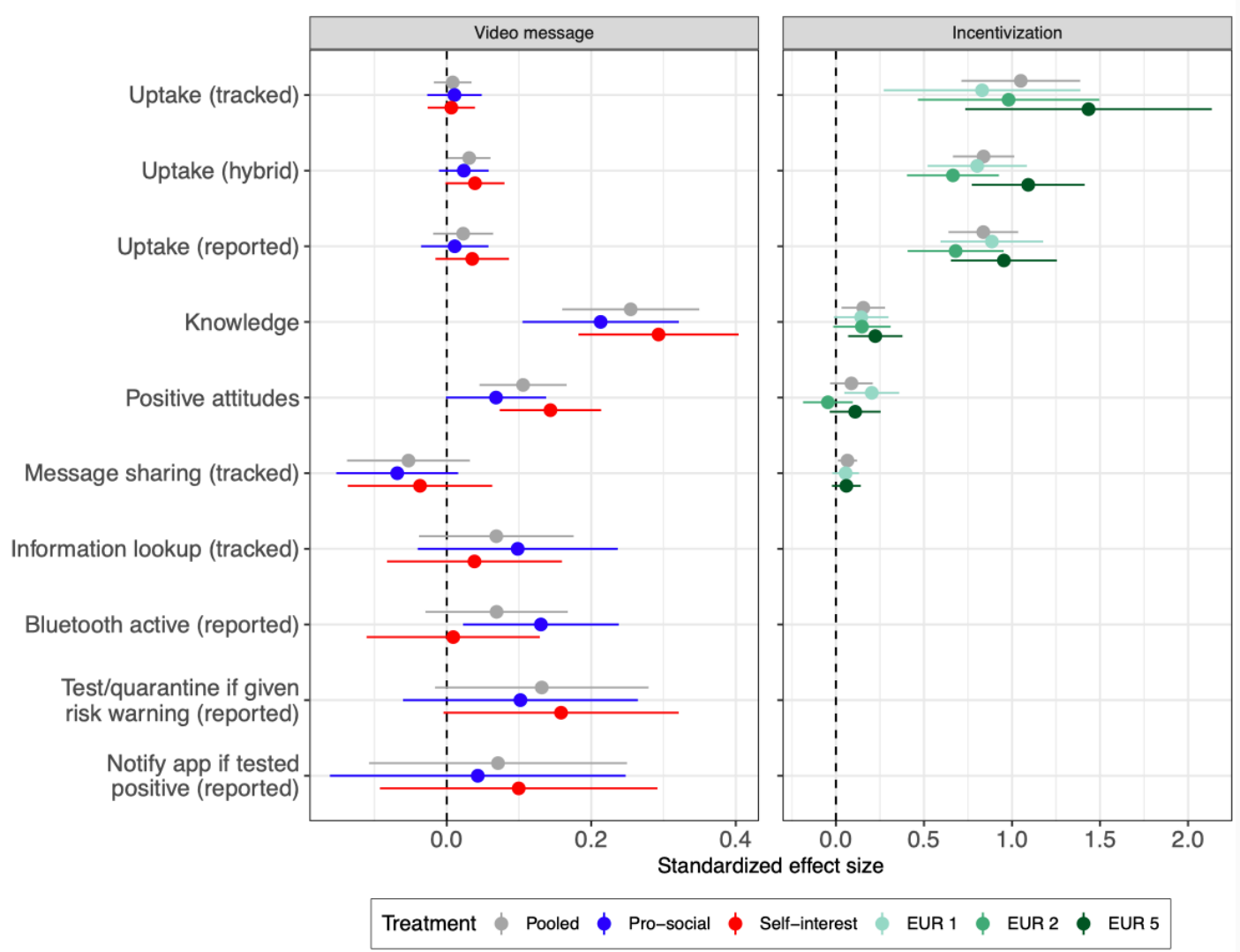
Hertie School

How can app usage be promoted?

- Information about app: goal, functionality, privacy
- Pro-social appeal: https://youtu.be/dyhDd_vrGEE
- Self-interest appeal: https://youtu.be/suOCvlW8_R0



Experimental results



Observational designs

Dilemmas of experimental research

- High internal validity at the cost of external validity
- Many of the actual treatments of interest cannot be (ethically) randomized (think: COVID, certain policies, election outcomes, ...)

(Some of the) problems with observational data

- Confounding
- Selection bias
- Limited generalizability of case studies

The silver lining

- The causal revolution: innovative research designs exploiting observational data to do causal inference
- ... often with the help of natural or quasi-experiments

Further reading

Hertie School