Building intelligent Global Health and
Humanitarian Response Technologies

BLOCK III, Week 10: *Trustworthy Evidence*

Context ***Plausible*** **≠ Proven**
Engineering ***Causal*** **Inference**

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

EPFL ARIADNE LABS ASHOKA UNIVERSITY WORLD ECONOMIC FORUM Rwanda Centre for the Fourth Industrial Revolution

If it makes sense and everyone believes it, why *test* it?

- Can a treatment make biological sense, be recommended in guidelines, and still be harmful?
- Are patients better off because we used it?

EPFL

Plausible ≠ Proven

"Obviously beneficial" Says who? For whom? Based on what?

Lecturer: *Prof. Mary-Anne Hartley*



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

CONTEXT

Outline

01 Plausible ≠ Proven

02 Association and confounding

03 Observational study designs

04 From association toward causation

05 Interventional studies and trials

CONTEXT

But it's OBVIOUS…isn't it?

Plausible treatments that became standard care

…until they were tested…

Examples:

1. Oxygen for premature babies
2. Steroids for head injury
3. Radical mastectomy

EXAMPLE 1

Oxygen for premature babies

CONTEXT

THE LOGIC

Premature babies have low oxygen saturation to underdeveloped lungs

Therefore: give more oxygen!

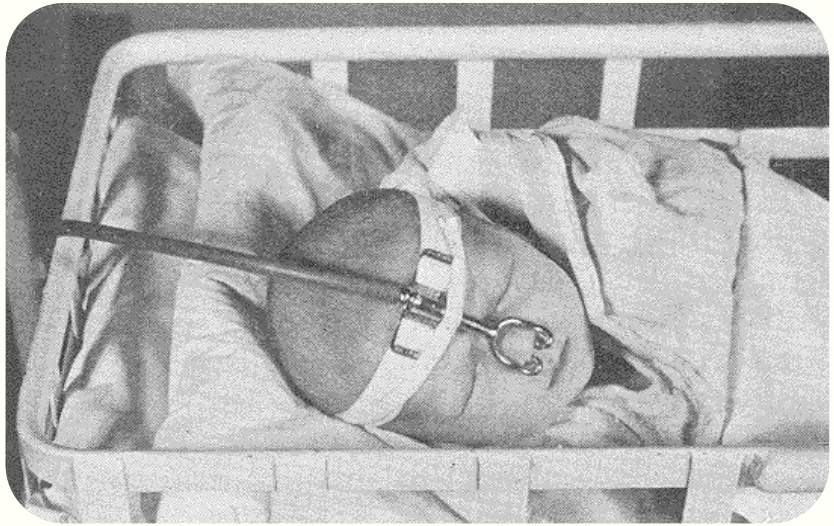


CONTEXT

Oxygen for premature babies

1940s

Generous oxygen in incubators = routine care





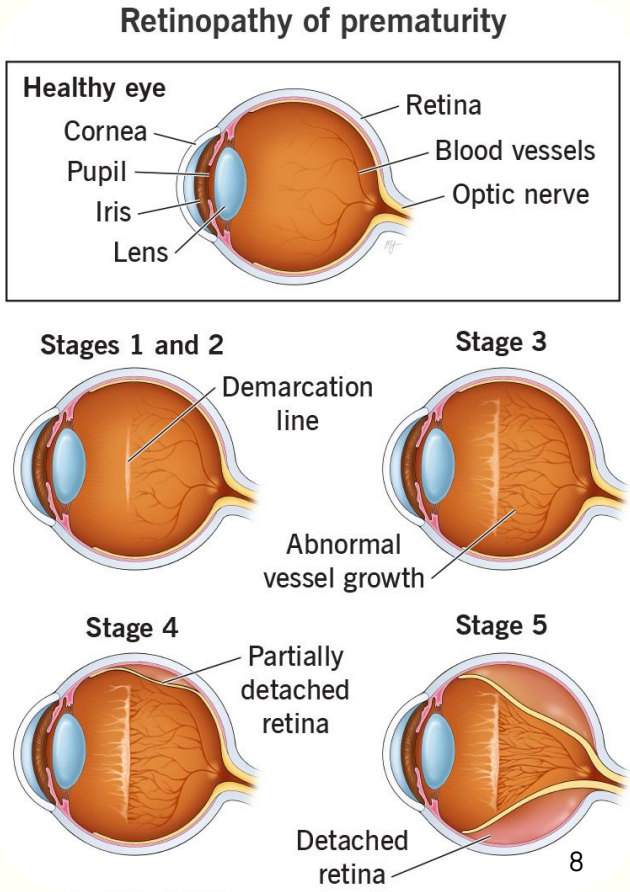
EXAMPLE 1

CONTEXT

Oxygen for premature babies

1940s — **Generous oxygen in incubators** = routine care

1942 — **New retinal disease** in premature babies reported



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

EXAMPLE 1

CONTEXT

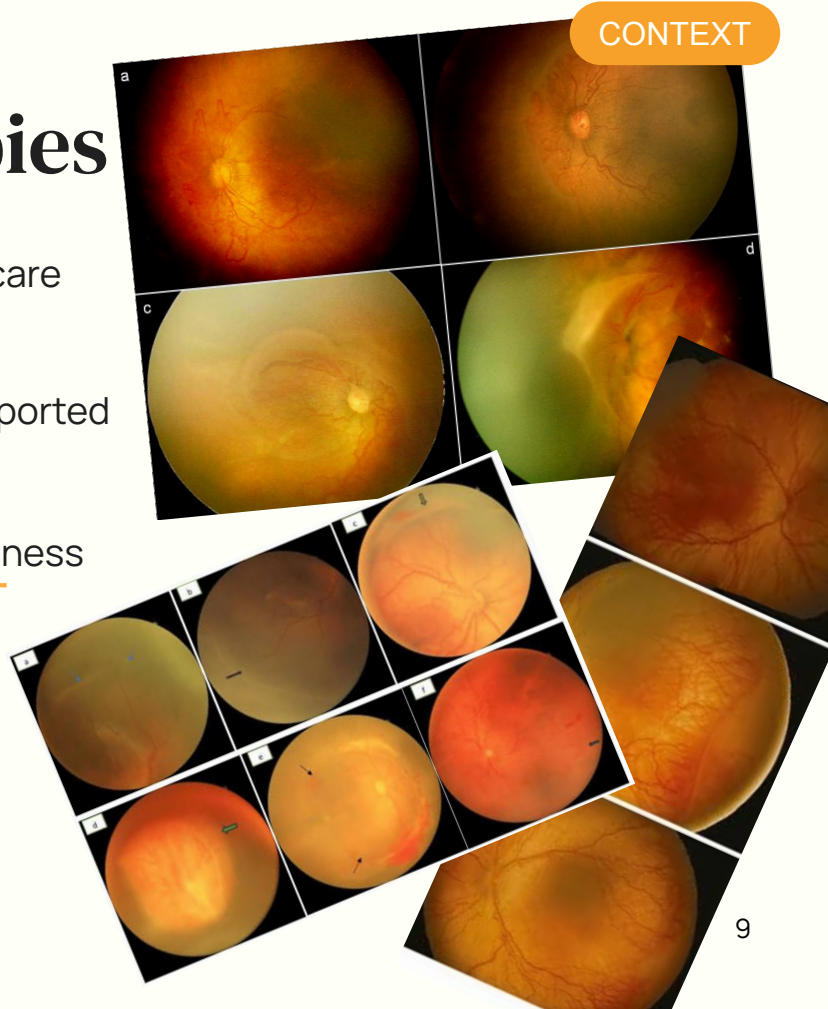
Oxygen for premature babies

1940s — **Generous oxygen in incubators** = routine care

1942 — **New retinal disease** in premature babies reported

1950 — Thousands of babies with unexplained blindness

PERMANENT



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

CONTEXT

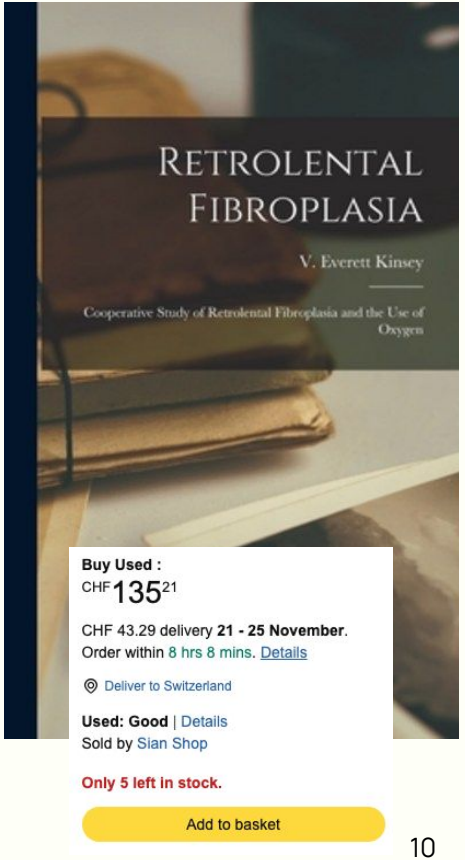
Oxygen for premature babies

1940s — **Generous oxygen in incubators** = routine care

1942 — **New retinal disease** in premature babies reported

1950 — Thousands of babies with unexplained blindness

1953 — **RANDOMIZED CONTROLLED TRIAL**
Hypothesis: oxygen is the cause



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Kinsey VE. Arch Ophthalmol 1956;56:481–543. Design as described in Silverman WA, 1980, chapter 6.

CONTEXT

How the trial worked

STAGE 1: RANDOMIZED 1 : 2

68 babies
CONTROL
routine oxygen

136 babies
INTERVENTION
restricted oxygen

Kinsey VE. Arch Ophthalmol 1956;56:481–543. Design as described in Silverman WA, 1980, chapter 6.

CONTEXT

How the trial worked

STAGE 1: RANDOMIZED 1 : 2

68 babies
CONTROL
routine oxygen

136 babies
INTERVENTION
restricted oxygen

Preplanned safety check: compare deaths only

No mortality difference detected → NEXT STAGE

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Kinsey VE. Arch Ophthalmol 1956;56:481–543. Design as described in Silverman WA, 1980, chapter 6.

CONTEXT

How the trial worked

STAGE 1: RANDOMIZED 1 : 2

68 babies
CONTROL
routine oxygen

136 babies
INTERVENTION
restricted oxygen

Preplanned safety check: compare deaths only

No mortality difference detected → NEXT STAGE

STAGE 2: ALL NEW BABIES GET RESTRICTED OXYGEN

+ 580 restricted oxygen

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Kinsey VE. Arch Ophthalmol 1956;56:481–543. Design as described in Silverman WA, 1980, chapter 6.

CONTEXT

RESULTS

COMPARE EYE DISEASE: ONLY BABIES WITH ADEQUATE EYE FOLLOW-UP (ALIVE AT 40 DAYS AND EXAMINED)



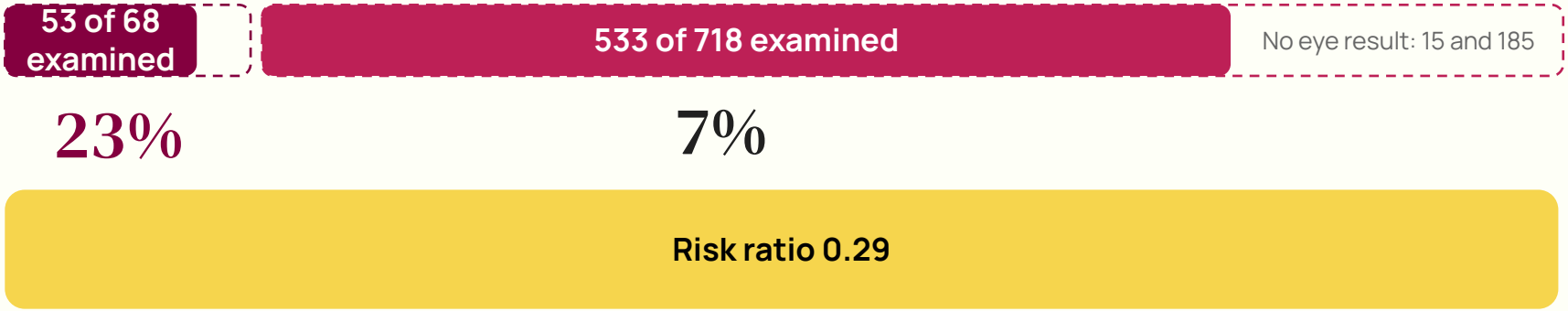
23%

7%

CONTEXT

RESULTS

COMPARE EYE DISEASE: ONLY BABIES WITH ADEQUATE EYE FOLLOW-UP (ALIVE AT 40 DAYS AND EXAMINED)



ROUTINE CARE caused the blindness

EXAMPLE 3

Radical mastectomy

CONTEXT

THE LOGIC

Breast cancer
spreads outward from the tumour

Therefore:
remove whole breast!



CONTEXT

Radical mastectomy

RANDOMIZED CONTROLLED TRIAL

NSABP B-04, 1971–74, multicentre, 1,665 women



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Fisher B et al. N Engl J Med 2002;347:567–575.

CONTEXT

Radical mastectomy

RANDOMIZED CONTROLLED TRIAL

NSABP B-04, 1971–74, multicentre, 1,665 women

SURVIVAL AT 25 YEARS

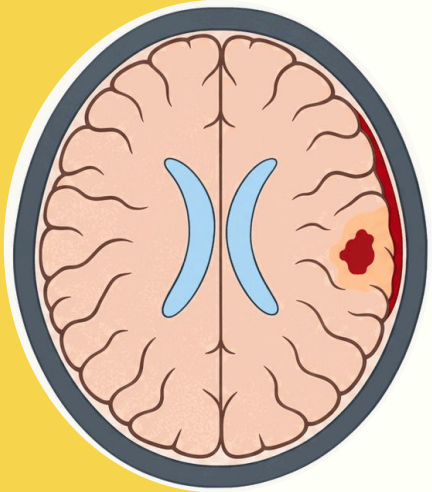
No demonstrated survival advantage

Radical mastectomy vs less extensive surgery

Fisher B et al. N Engl J Med 2002;347:567–575.

EXAMPLE 2

Steroids for head injury



CONTEXT

THE LOGIC

Head injury = brain swelling
Steroids = reduce swelling

Therefore:
give steroids in head injury!



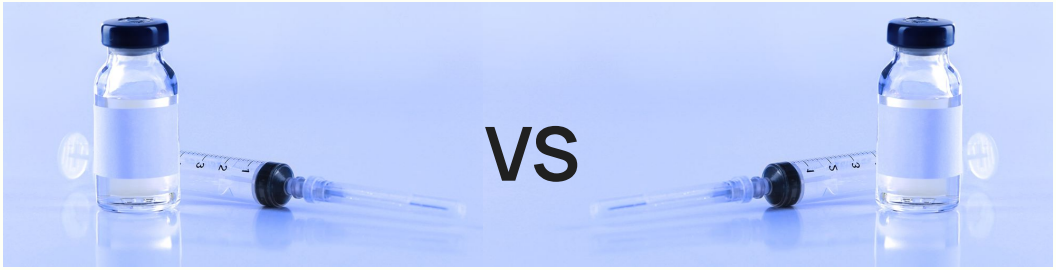
EXAMPLE 2

Steroids for head injury

RANDOMIZED PLACEBO-CONTROLLED TRIAL

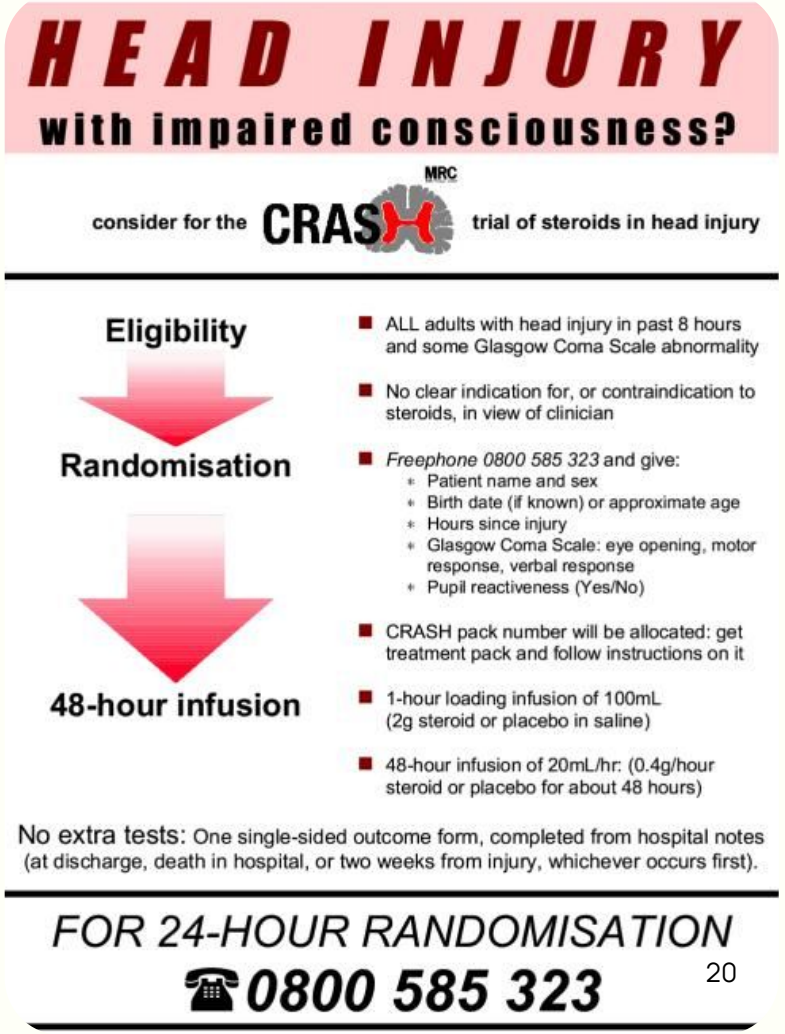
CRASH, 2004

- 10,008 patients
- 49 countries



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

CRASH trial collaborators. Lancet 2004;364:1321–1328.



CONTEXT

Steroids for head injury

DEATHS AT 2 WEEKS

21% steroids

18% placebo

Risk ratio 1.18 (95% CI 1.09–1.27)
1.18-fold more deaths

STOP

Stopped early

	Dead		
	Corticosteroid (n=5007)	Placebo (n=5001)	Relative risk (CI)
Severity of injury			
Severe (GCS 3–8)	912/1937 (47·1%)	808/1914 (42·2%)	
Moderate (GCS 9–12)	247/1497 (16·5%)	181/1422 (12·7%)	



CRASH trial collaborators. Lancet 2004;364:1321–1328.

Finding the cause

1 Observe

Watch what happens “in the wild”

1.1 Case series

1.2 Cross-sectional study

1.3 Ecological study

1.4 Case-control study

1.5 Cohort study

2 Intervene

Assign the treatment on purpose, then compare.

2.1 Before-and-after study

2.2 Controlled trial

2.3 Randomised trial

Finding the cause

1 Observe

Watch what happens "in the wild"

1.1 Case series

1.2 Cross-sectional study

1.3 Ecological study

1.4 Case-control study

1.5 Cohort study

2 Intervene

Assign the treatment on purpose, then compare.

2.1 Before-and-after study

2.2 Controlled trial

2.3 Randomised trial

CONTEXT

1 Observe

2 ways to observe

DESCRIBE

What is happening, to whom, where and when?

1.1 Case series
1.2 Cross-sectional study
1.3 Ecological study

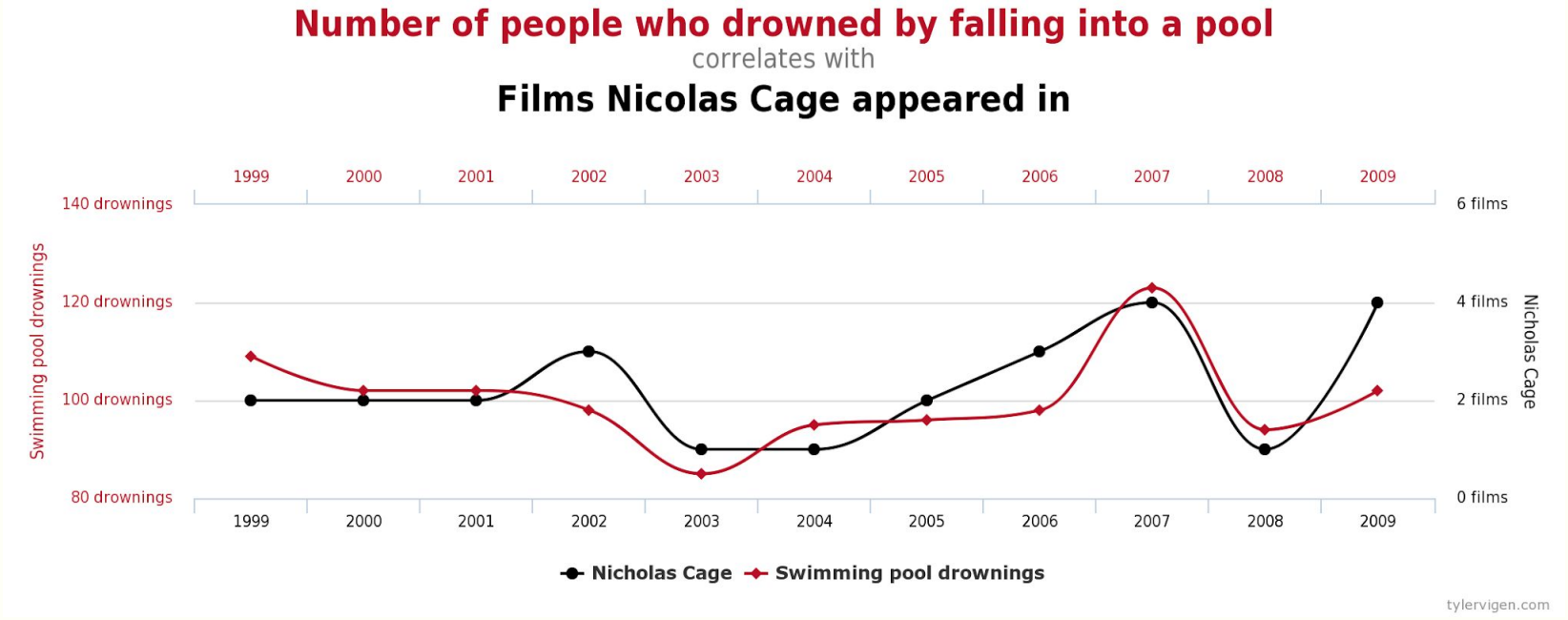
COMPARE

Is it more common in one group than another?

1.4 Case-control study
1.5 Cohort study

CONTEXT

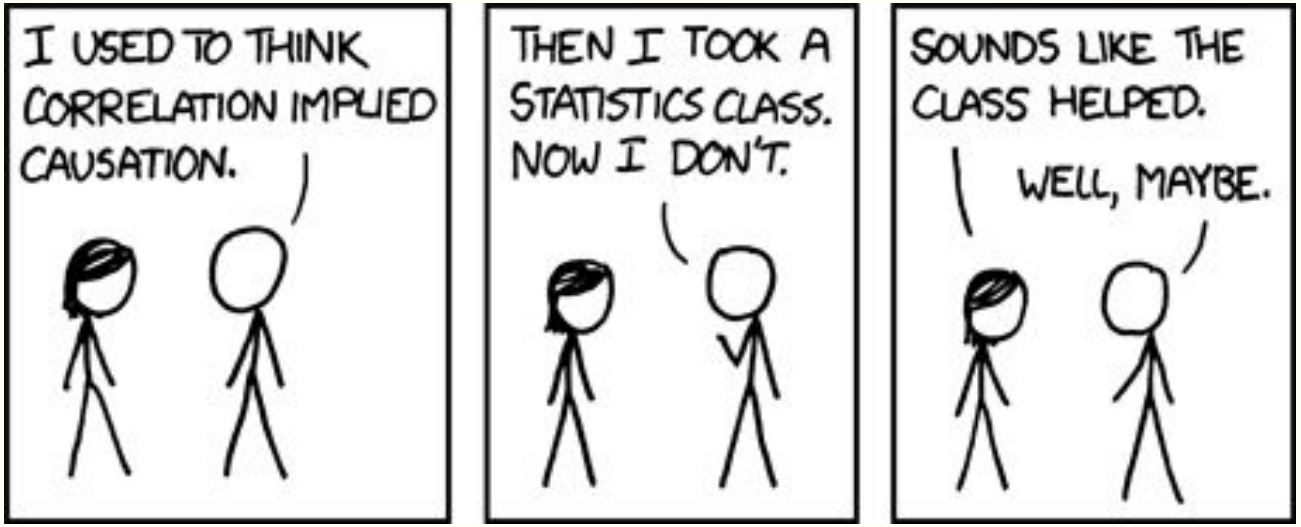
Association ≠ causation



Tyler Vigen, original chart, CC BY 4.0

CONTEXT

Correlation ≠ causation



Randall Munroe, xkcd #552, "Correlation", CC BY-NC 2.5

CONTEXT

Association or causation?

ASSOCIATION

Two things
happen together

CAUSATION

Change one
and the other changes

Seeing is not doing

CONTEXT

Do shoes cause headaches?

THE OBSERVATION

People who sleep with their shoes on wake up with a headache



Therefore: shoes cause headaches!



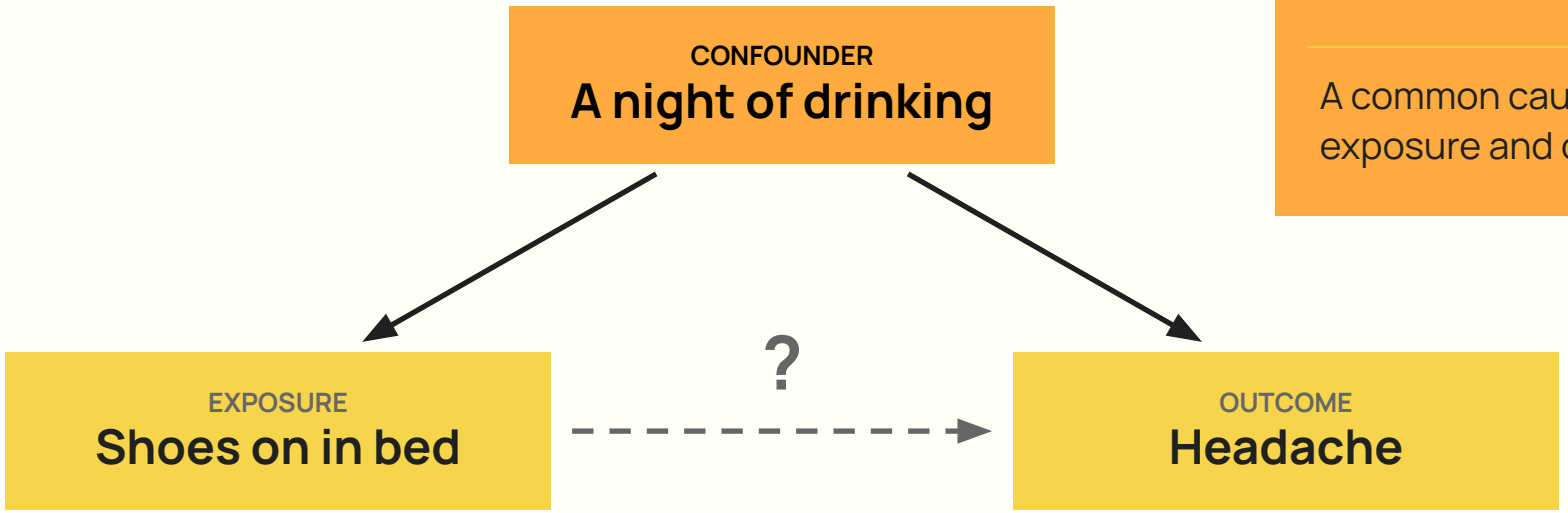
CONTEXT

Confounding

Terminology

Confounder

A common cause of exposure and outcome.



CONTEXT

Which treatment is better?

Success	Open surgery	Percutaneous
All stones	78%	**83%**

Which would you choose?

Charig et al. BMJ 1986;292:879–882; Pearl & Mackenzie, chapter 6.

CONTEXT

Which treatment is better?

Success	Open surgery	Percutaneous
All stones	78%	**83%**
Small stones	**93%**	87%
Large stones	**73%**	69%

Open surgery wins for small stones and for large stones. It loses overall.

QUESTION 1

What is the confounder?

QUESTION 2

What is this situation called?

Charig et al. BMJ 1986;292:879–882; Pearl & Mackenzie, chapter 6.

CONTEXT

Which treatment is better?

Success	Open surgery	Percutaneous
All stones	78%	**83%**
Small stones	**93%**	87%
Large stones	**73%**	69%

Open surgery wins for small stones and for large stones. It loses overall.

THE CONFOUNDER

Stone size

Large stones are harder to treat, and most of them went to open surgery.

THE SITUATION

Simpson's paradox

The pattern within each subgroup reverses when the groups are combined.

Charig et al. BMJ 1986;292:879–882; Pearl & Mackenzie, chapter 6.

CONTEXT

1 Observe

Watch what happens “in the wild”

1.1 Case series

1.2 Cross-sectional study

1.3 Ecological study

1.4 Case-control study

1.5 Cohort study

Notice a pattern. Raise a question.

CONTEXT

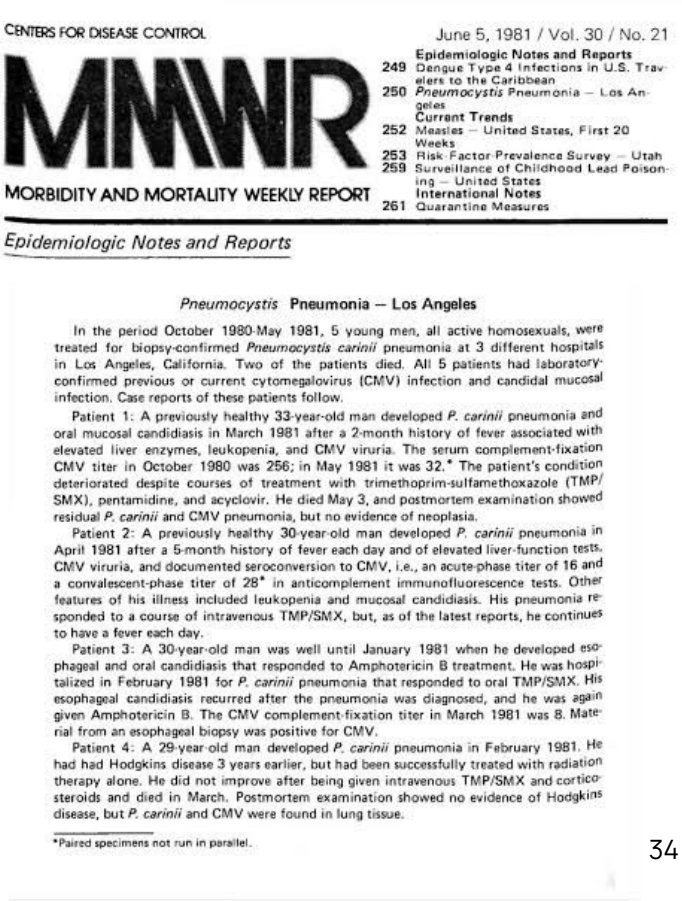
Good example

LOS ANGELES, 1981

Five young men with a rare pneumonia

Therefore:

something new is happening…



CENTERS FOR DISEASE CONTROL

MMWR

MORBIDITY AND MORTALITY WEEKLY REPORT

June 5, 1981 / Vol. 30 / No. 21

Epidemiologic Notes and Reports
249 Dengue Type 4 Infections in U.S. Travelers to the Caribbean
250 *Pneumocystis* Pneumonia — Los Angeles
Current Trends
252 Measles — United States, First 20 Weeks
253 Risk-Factor-Prevalence Survey — Utah
259 Surveillance of Childhood Lead Poisoning — United States
International Notes
261 Quarantine Measures

Epidemiologic Notes and Reports

Pneumocystis Pneumonia — Los Angeles

In the period October 1980-May 1981, 5 young men, all active homosexuals, were treated for biopsy-confirmed *Pneumocystis carinii* pneumonia at 3 different hospitals in Los Angeles, California. Two of the patients died. All 5 patients had laboratory-confirmed previous or current cytomegalovirus (CMV) infection and candidal mucosal infection. Case reports of these patients follow.

Patient 1: A previously healthy 33-year-old man developed *P. carinii* pneumonia and oral mucosal candidiasis in March 1981 after a 2-month history of fever associated with elevated liver enzymes, leukopenia, and CMV viruria. The serum complement-fixation CMV titer in October 1980 was 256; in May 1981 it was 32.* The patient's condition deteriorated despite courses of treatment with trimethoprim-sulfamethoxazole (TMP/SMX), pentamidine, and acyclovir. He died May 3, and postmortem examination showed residual *P. carinii* and CMV pneumonia, but no evidence of neoplasia.

Patient 2: A previously healthy 30-year-old man developed *P. carinii* pneumonia in April 1981 after a 5-month history of fever each day and of elevated liver-function tests. CMV viruria, and documented seroconversion to CMV, i.e., an acute-phase titer of 16 and a convalescent-phase titer of 28* in anticomplement immunofluorescence tests. Other features of his illness included leukopenia and mucosal candidiasis. His pneumonia responded to a course of intravenous TMP/SMX, but, as of the latest reports, he continues to have a fever each day.

Patient 3: A 30-year-old man was well until January 1981 when he developed esophageal and oral candidiasis that responded to Amphotericin B treatment. He was hospitalized in February 1981 for *P. carinii* pneumonia that responded to oral TMP/SMX. His esophageal candidiasis recurred after the pneumonia was diagnosed, and he was again given Amphotericin B. The CMV complement-fixation titer in March 1981 was 8. Material from an esophageal biopsy was positive for CMV.

Patient 4: A 29-year-old man developed *P. carinii* pneumonia in February 1981. He had had Hodgkins disease 3 years earlier, but had been successfully treated with radiation therapy alone. He did not improve after being given intravenous TMP/SMX and corticosteroids and died in March. Postmortem examination showed no evidence of Hodgkins disease, but *P. carinii* and CMV were found in lung tissue.

*Paired specimens not run in parallel.

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

CDC. Pneumocystis pneumonia, Los Angeles. MMWR 1981;30:250–252.

CONTEXT

Bad example

12 children with bowel symptoms and developmental regression
8/12 parents linked it to the MMR vaccine

Therefore:
vaccines cause autism…?



EARLY REPORT

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associ… by the parents, with measles, mumps, and rub… vaccination in eight of the 12 children, with measl… infection in one child, and otitis media in an… . All 1… children had intestinal abnormalities ranging from lymphoid nodular hyperplasia to aphthoid ulceration. Histology showed patchy chronic inflammation … in 11 children and reactive ileal lymphoid hyperplasia in seven, but no granulomas. Be…oural disor… included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (one). There were no focal neurological abnormalities and MRI and EEG tests were normal. Abnor… laboratory results were significantly raised urinary …ethylmalonic acid compared with age-matched control… (p=0·003), low haemoglobin in four children, and low s… IgA in four children.

Interpretation We identi… associated gastrointestinal dis…se and developmental regression in a group of previ… normal c…n, which was generally associated in time … possible environmental triggers.

Lancet 199…**351:** 637–41

Introduction

We saw several children who, after a p… of apparent normality, lost acquired skills, includ… communication. They all had gastrointestinal …mptoms, including abdominal pain, diarrhoea, and …ating and, in some cases, food intolerance. We …scribe… clinical f…dings, and gastrointestinal featur… of these chi…en.

Patients and meth…

12 children, con…cutively …ferred to … department of paediatric gastr…terology … a history of a pervasive developmental …der with loss …ired skills and intestinal symptoms (…arrhoea, abdominal …in, bloating and food intolerance), were inv…gated. All children were admitted to the ward for a week, accomp…ed by their parents.

Clinical investigations

… took histor… including details of immunisations and …posure to infectious diseases, and assessed the children. In 11 ca… the history was obtained by the senior clinician (JW-S). Neur… and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.[1] Developmental … included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and …

RETRACTED

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Wakefield AJ et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998;351:637–641. RETRACTED: Lancet 2010;375:445.

CONTEXT

Limitation of a case series

No control group

Where it failed

- Only 12 children
- No comparison group
- Was autism more common after vaccination?

Larger cohort studies found no link.



EARLY REPORT

Early report

Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

A J Wakefield, S H Murch, A Anthony, J Linnell, D M Casson, M Malik, M Berelowitz, A P Dhillon, M A Thomson, P Harvey, A Valentine, S E Davies, J A Walker-Smith

Summary

Background We investigated a consecutive series of children with chronic enterocolitis and regressive developmental disorder.

Methods 12 children (mean age 6 years [range 3–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Ileocolonoscopy and biopsy sampling, magnetic-resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

Findings Onset of behavioural symptoms was associ... by the parents, with measles, mumps, and rub... vaccination in eight of the 12 children, with measl... infection in one child, and otitis media in an... All 1... children had intestinal abnormalities ranging from lymphoid nodular hyperplasia to aphthoid ulceration. Histology showed patchy chronic inflammation in 11 children and reactive ileal lymphonodular hyperplasia in seven, but no granulomas. Behavioural disorders included autism (nine), disintegrative psychosis (one), and possible postviral or vaccinal encephalitis (two). There were no focal neurological abnormalities and CSF and EEG tests were normal. Abnormal laboratory results were significantly raised urinary methylmalonic acid compared with age-matched controls (p=0·003), low haemoglobin in four children, and low serum IgA in four children.

Interpretation We identified associated gastrointestinal disease and developmental regression in a group of previously normal children, which was generally associated in time with possible environmental triggers.

Lancet 199... **351:** 637–41

Introduction

We saw several children who, after a period of apparent normality, lost acquired skills, including communication. They all had gastrointestinal symptoms, including abdominal pain, diarrhoea, and bloating and, in some cases, food intolerance. We describe the clinical findings, and gastrointestinal features of these children.

Patients and methods

12 children, consecutively referred to the department of paediatric gastroenterology with a history of a pervasive developmental disorder with loss of acquired skills and intestinal symptoms (diarrhoea, abdominal pain, bloating and food intolerance), were investigated. All children were admitted to the ward for 1 week, accompanied by their parents.

Clinical investigations

We took histories, including details of immunisations and exposure to infectious diseases, and assessed the children. In 11 cases the history was obtained by the senior clinician (JW-S). Neurological and psychiatric assessments were done by consultant staff (PH, MB) with HMS-4 criteria.[1] Developmental assessment included a review of prospective developmental records from parents, health visitors, and general practitioners. Four children did not undergo psychiatric assessment in hospital; all had been assessed professionally elsewhere, so these assessments were used as the basis for their behavioural diagnosis.

After bowel preparation, ileocolonoscopy was performed by SHM or MAT under sedation with midazolam and pethidine. Paired frozen and formalin-fixed mucosal biopsy samples were taken from the terminal ileum; ascending, transverse, descending, and sigmoid colons, and from the rectum. The procedure was recorded by video or still images, and were compared with images of the previous seven consecutive paediatric colonoscopies (four normal colonoscopies and three on children with ulcerative colitis), in which the physician reported normal appearances in the terminal ileum. Barium follow-through radiography was possible in some cases.

Also under sedation, cerebral magnetic-resonance imaging (MRI), electroencephalography (EEG) including visual, brain stem auditory, and sensory evoked potentials (where compliance made these possible), and lumbar puncture were done.

Laboratory investigations

Thyroid function, serum long-chain fatty acids, and

RETRACTED

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Wakefield AJ et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998;351:637–641. RETRACTED: Lancet 2010;375:445.

CONTEXT

1 Observe

Watch what happens "in the wild"

1.1 Case series

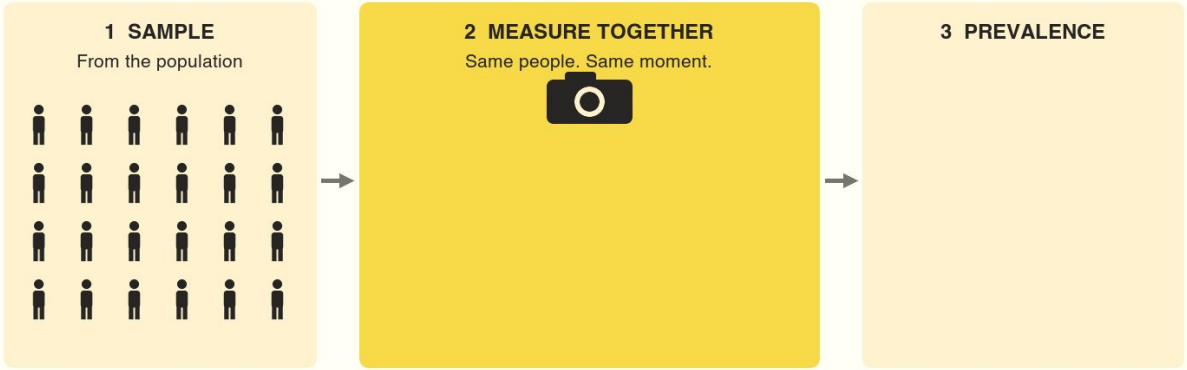
1.2 Cross-sectional study

1.3 Ecological study

1.4 Case-control study

1.5 Cohort study

A sample. A snapshot. Prevalence.



No follow-up: we see who has what now, not which came first.

CONTEXT

Good example

GENEVA, APRIL TO MAY 2020

Blood samples from 2,766 residents, taken once
About 1 in 10 already had antibodies to SARS-CoV-2

Therefore: most infections go uncounted

Prevalence existing cases ÷ population, at one point in time

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Stringhini S et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP). Lancet 2020;396:313–319.

CONTEXT

Bad example

One survey, one day
People on a diet weigh more than people who are not

Therefore: dieting makes you heavier?

Illustrative example.

CONTEXT

Limitation of a cross-sectional study

Unclear temporal order

Where it failed

Dieters weighed more on the survey day. Reading that as evidence that dieting caused weight gain overlooks the reverse sequence: weight gain could have prompted the diet.

Illustrative example.

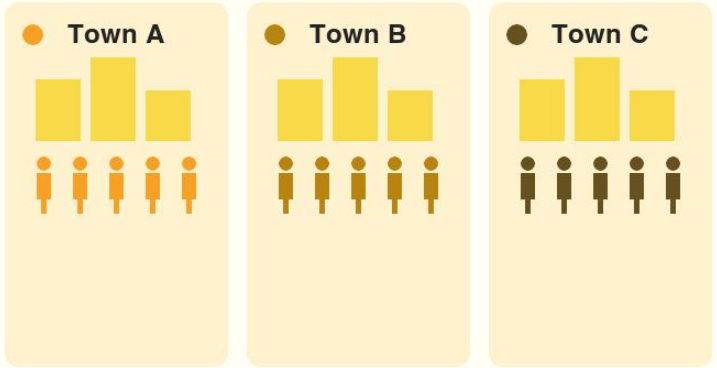
CONTEXT

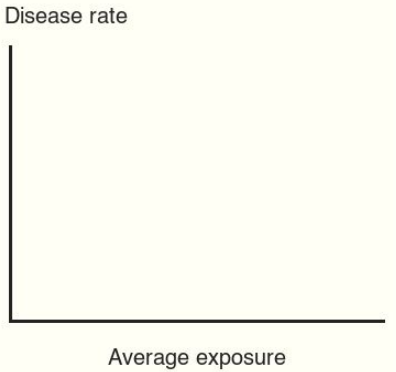
1 Observe

Watch what happens “in the wild”

1.1 Case series
1.2 Cross-sectional study
1.3 Ecological study
1.4 Case-control study
1.5 Cohort study

Compare groups, not individuals.





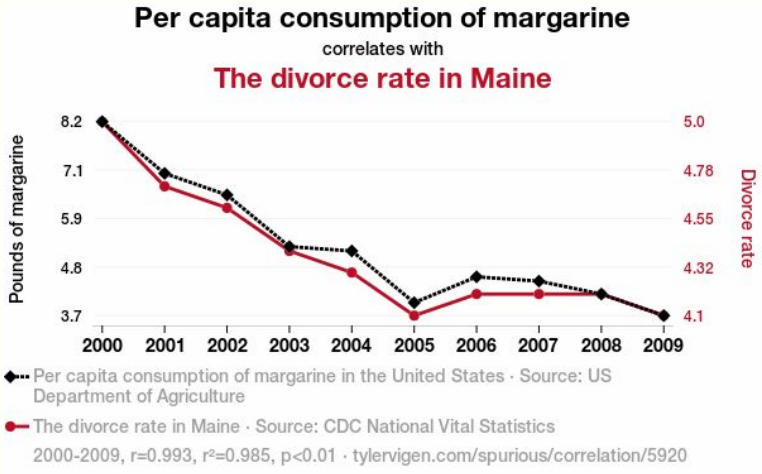
CONTEXT

Example: margarine and divorce

MAINE, 2000 TO 2009

Americans eat less margarine
and the divorce rate in Maine falls with it

Therefore: margarine ends marriages?



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Tyler Vigen, original aggregate-data example, CC BY 4.0. Not evidence of a causal effect.

CONTEXT

Where it fails

Countries that eat more chocolate win more Nobel prizes
A real correlation, across 23 countries

Therefore: eat chocolate, win a Nobel?

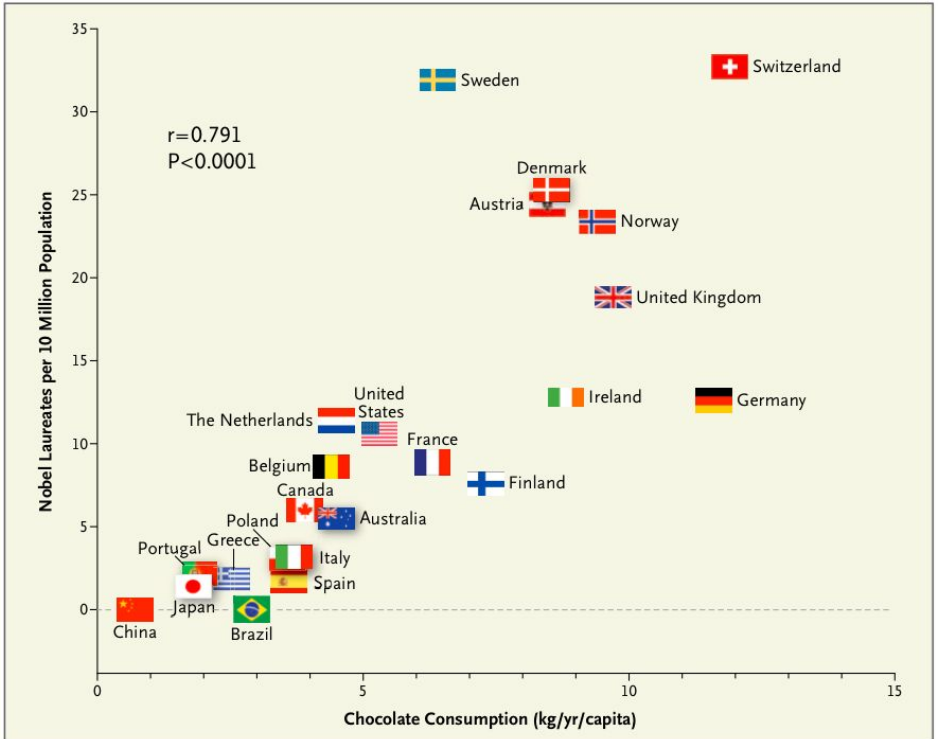


Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Messerli FH. Chocolate consumption, cognitive function, and Nobel laureates. N Engl J Med 2012;367:1562–1564.

CONTEXT

Limitation of an ecological study

Ecological fallacy

Where it failed

The researchers never measured what the laureates ate, so the country-level pattern cannot support advice to eat chocolate.

Terminology

Ecological fallacy

Judging individuals from the average of their group.

Messerli FH. Chocolate consumption, cognitive function, and Nobel laureates. N Engl J Med 2012;367:1562–1564.

CONTEXT

1 Observe

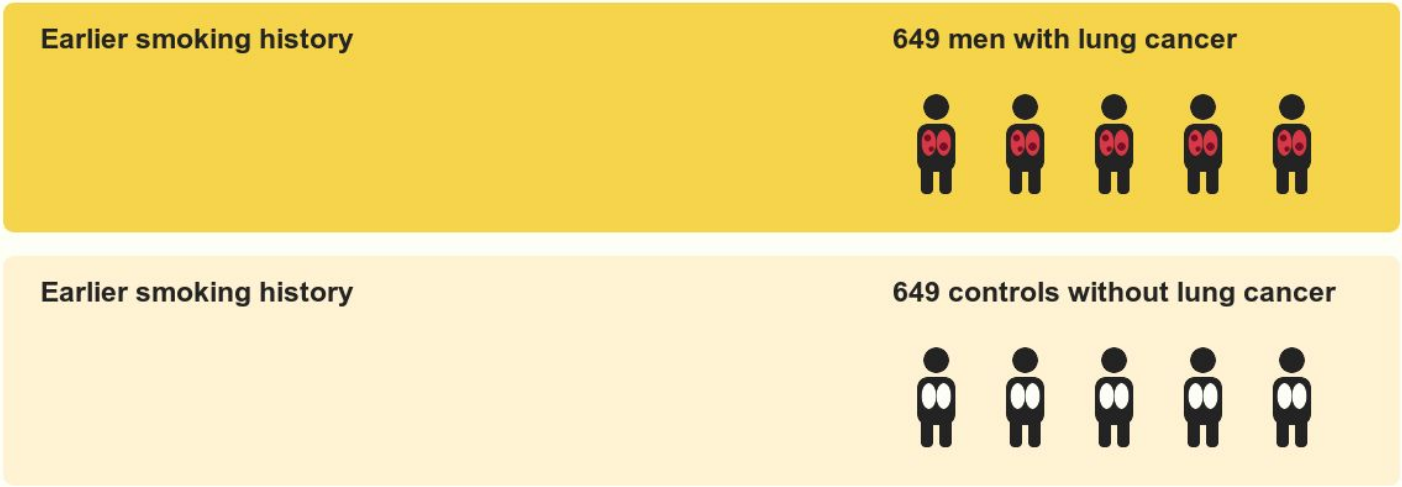
Watch what happens “in the wild”

1.1 Case series
1.2 Cross-sectional study
1.3 Ecological study
1.4 Case-control study
1.5 Cohort study



CONTEXT

Example: lung cancer vs never smoking



Odds ratio
exposure odds in cases ÷ exposure odds in controls

Doll R, Hill AB. Smoking and carcinoma of the lung. BMJ 1950;2:739–748.

CONTEXT

Where it fails

Ask people with food poisoning what they ate this week
They list every meal. The healthy controls cannot remember Tuesday's lunch

Therefore: every food looks guilty

Illustrative example.

CONTEXT

Limitation of a case-control study

Recall bias

Where it failed

After a food-poisoning outbreak, cases may recall a shared meal more readily than healthy controls.
That difference in memory can make the meal look like the source.

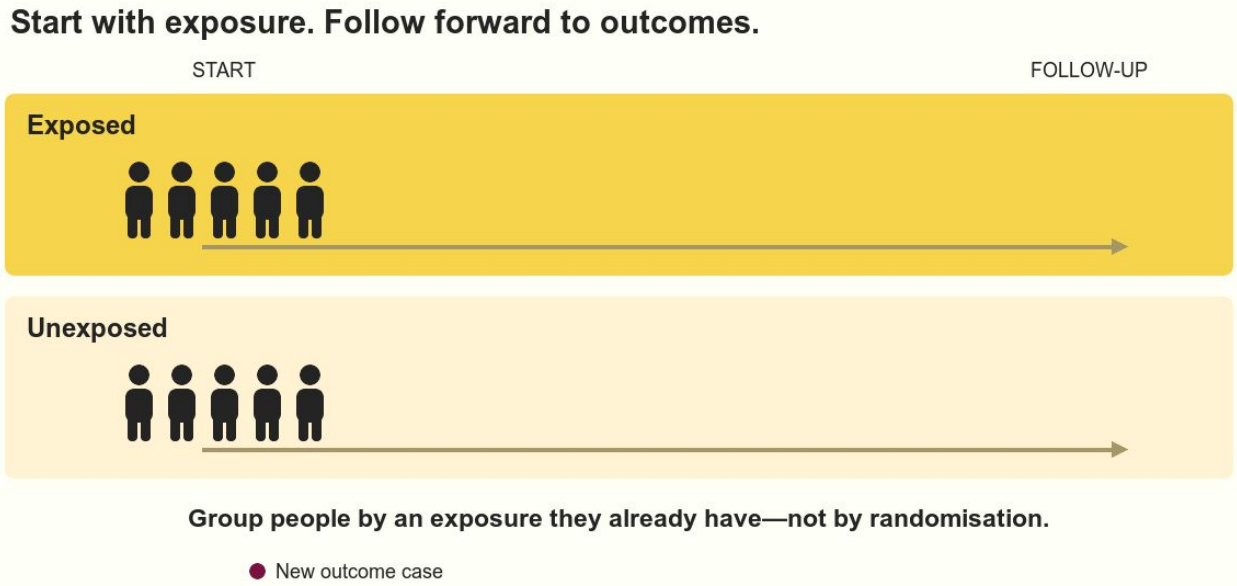
Illustrative example.

CONTEXT

1 Observe

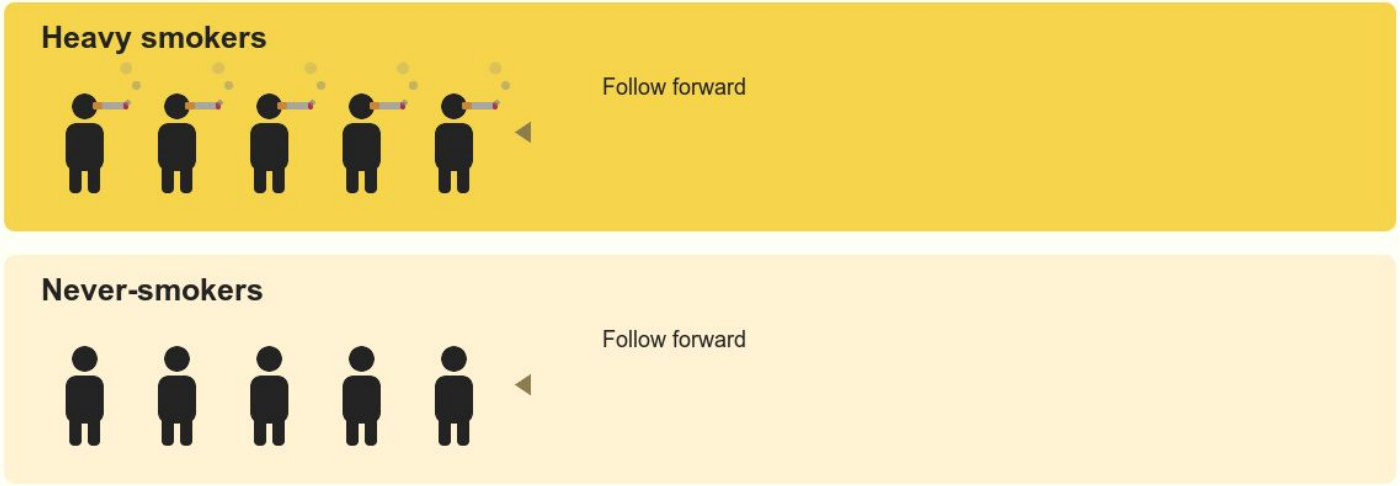
Watch what happens “in the wild”

1.1 Case series
1.2 Cross-sectional study
1.3 Ecological study
1.4 Case-control study
1.5 Cohort study



CONTEXT

Example: smoking MDs vs lung cancer deaths



Lung-cancer mortality
deaths per 1,000 person-years

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking. BMJ 1956;2:1071–1081.

CONTEXT

Where it fails

Cohort studies: women on hormone therapy had far less heart disease
So millions were prescribed it to protect their hearts

Therefore: hormones prevent heart attacks?

0.61
Observational cohort: relative risk of heart disease in hormone users

Grodstein F et al. N Engl J Med 1996;335:453–461. Writing Group for the Women's Health Initiative Investigators. JAMA 2002;288:321–333.

CONTEXT

Limitation of a cohort study

Confounding

Where it failed

Women who chose hormone therapy also differed in health and wealth.

In the randomised trial, **combined hormone therapy increased coronary events.**

0.61

Observational cohort: relative risk of heart disease in hormone users

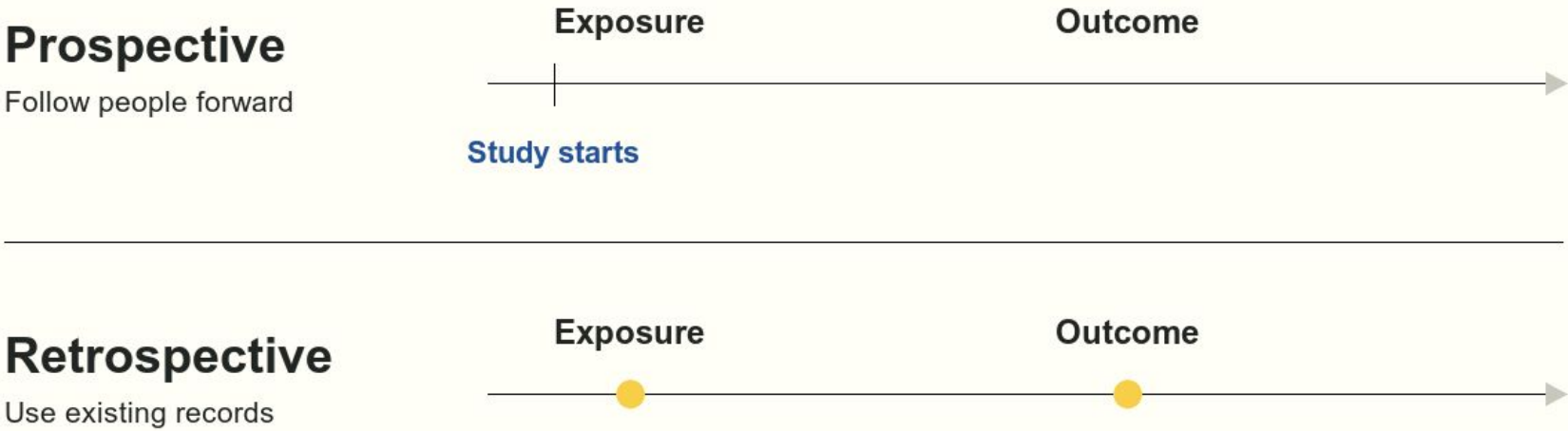
1.29

Randomised trial, 2002: hazard ratio for heart disease

Grodstein F et al. N Engl J Med 1996;335:453–461. Writing Group for the Women's Health Initiative Investigators. JAMA 2002;288:321–333.

CONTEXT

Prospective and retrospective cohorts



CONTEXT

Bradford Hill: nine considerations

01 Strength
How big is the effect?

02 Consistency
Seen again and again

03 Specificity
A specific exposure and outcome (often absent)

04 Temporality
Cause before effect

05 Biological gradient
More exposure, more disease

06 Plausibility
A mechanism that makes sense

07 Coherence
Fits what we already know

08 Experiment
Remove the cause, the effect falls

09 Analogy
Similar causes, similar effects

Hill. The environment and disease: association or causation? 1965.

CONTEXT

How many times? Relative effect sizes

Risk or rate ratio

rate in the exposed ÷ rate in the unexposed

Lung-cancer deaths per 1,000 a year: 1.66 in heavy smokers ÷ 0.07 in never-smokers = 24

Odds ratio (OR)

odds of exposure in cases ÷ odds in controls

Doll and Hill, 1950: about 14

Relative risk reduction

1 – risk ratio

Salk vaccine: about 71% fewer cases in one polio season

No difference means a ratio of 1 or a relative risk reduction of 0%.

Cochrane Handbook, chapter 6. Magnitude is not proof of causation.

CONTEXT

How much difference? Absolute sizes

Risk or rate difference	Number needed to treat (NNT)	Always state the period
rate in one group − rate in the other	1 ÷ risk difference	a risk, a risk difference and an NNT belong to a follow-up time
1.66 − 0.07 = 1.6 extra lung-cancer deaths per 1,000 men a year	Salk vaccine: about 2,500 children vaccinated to prevent one case in one polio season	Salk trial: one polio season, 1954

Same trial, two headlines: 71% fewer cases, or 4 fewer per 10,000 children.

Cochrane Handbook, chapter 6. Magnitude is not proof of causation.

CONTEXT

4. Temporality: exposure precedes outcome

Temporality, Latency, Reverse causation

Smoking begins → Lung cancer develops

Time between exposure and disease = latency

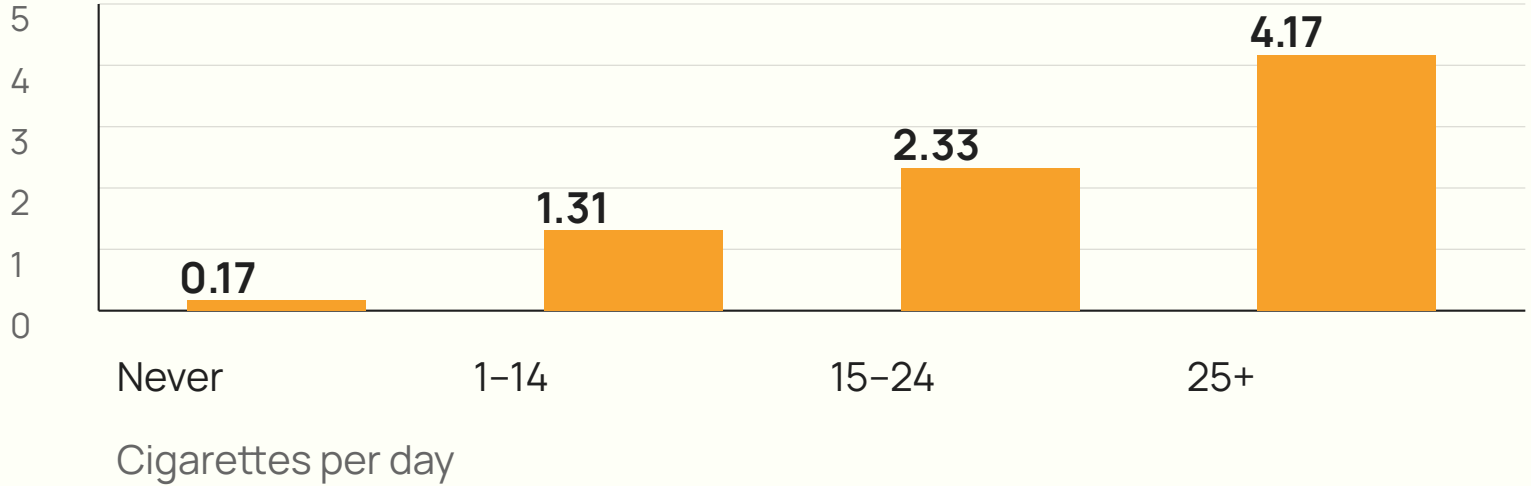
Use dates, repeated measurements and a biologically credible lag. There is no universal temporality score.

Hill, 1965; British Doctors Study. Temporal order is necessary, not sufficient.

CONTEXT

5. Biological gradient

Y-axis: annual lung-cancer deaths per 1,000 men



Doll et al., BMJ 2004, Table 1. Age/year-standardised rates; not percent or lifetime risk.

Finding the cause

1 Observe

Watch what happens "in the wild"

1.1 Case series
1.2 Cross-sectional study
1.3 Ecological study

1.4 Case-control study
1.5 Cohort study

2 Intervene

Assign the treatment on purpose, then compare.

2.1 Before-and-after study
2.2 Controlled trial
2.3 Randomised trial

CONTEXT

The counterfactual

Terminology

Counterfactual

Outcome under alternative care.

Treatment effect: difference between outcomes under the two care strategies.

AI-generated teaching image

CONTEXT

2 Intervene

Assign the treatment on purpose, then compare.

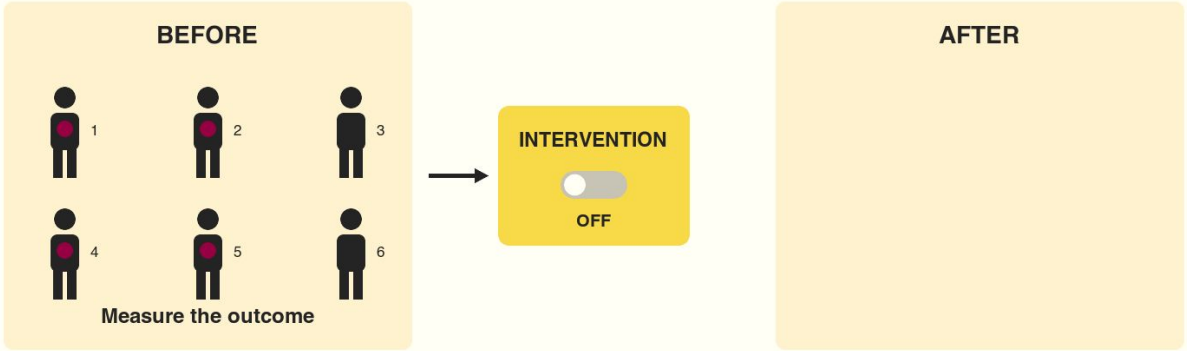
2.1 Before-and-after study

2.2 Controlled trial

2.3 Randomised trial

BEFORE-AND-AFTER STUDY

The same group, measured twice.



No separate control group

CONTEXT

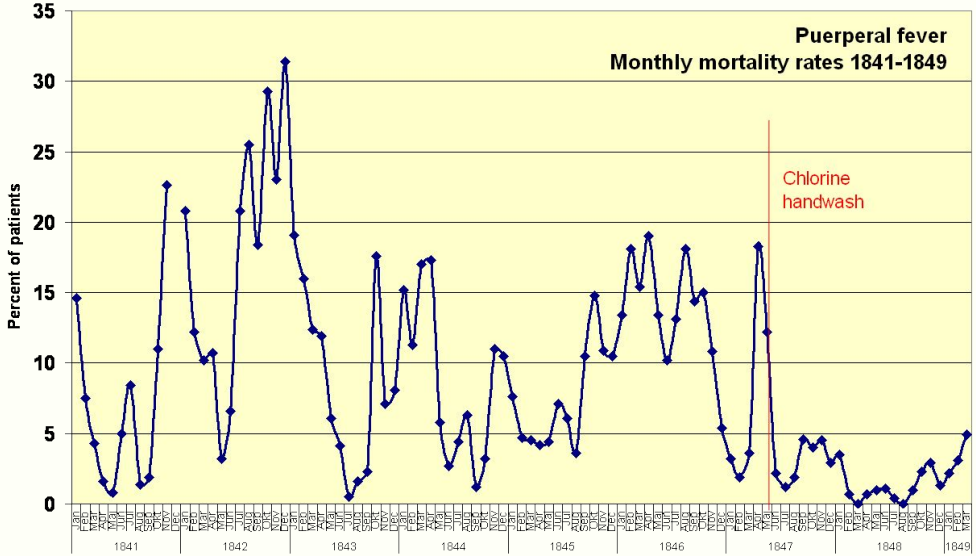
A before and after study that worked

VIENNA, 1847

Doctors start washing their hands in chlorine

and deaths on the maternity ward fall from 18% to 2%

Therefore: handwashing saves lives



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Image: monthly mortality, Vienna maternity clinic, 1841 to 1849, via Wikimedia Commons.

CONTEXT

Where it fails

Before: no umbrella, and it rained every day

After: I bought an umbrella, and it has not rained since

Therefore: my umbrella stops the rain?

Illustrative example.

CONTEXT

Limitation of a before-and-after study

No concurrent control group

Where it failed

The rain also stopped for my neighbour who never bought one. A before-and-after comparison would give my umbrella credit for the weather.

Illustrative example.

CONTEXT

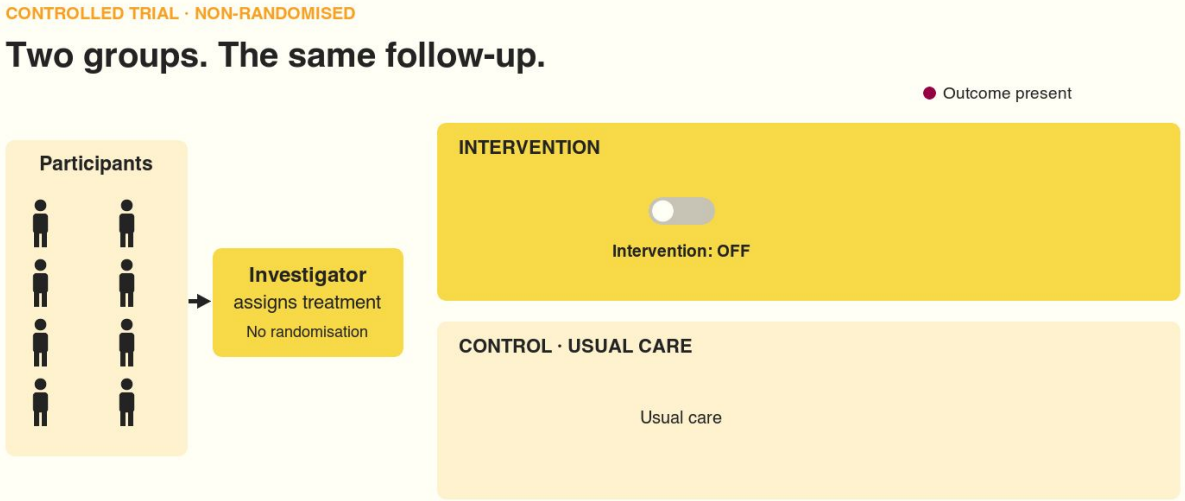
2 Intervene

Assign the treatment on purpose, then compare.

2.1 Before-and-after study

2.2 Controlled trial

2.3 Randomised trial



CONTEXT

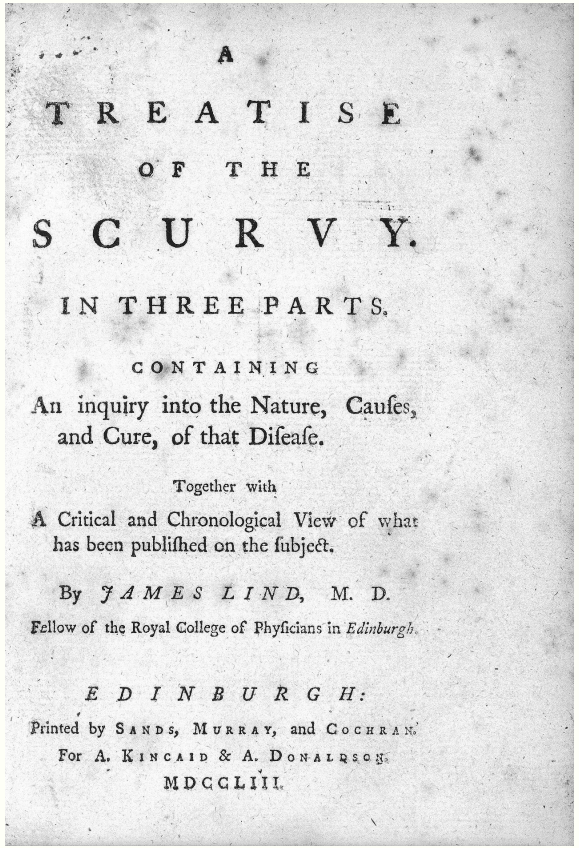
Example: scurvy at sea

HMS SALISBURY, 1747

12 sailors with scurvy, six pairs, six treatments

only the pair given oranges and lemons recovered

Therefore: citrus cures scurvy


A
TREATISE
OF THE
SCURVY.
IN THREE PARTS.
CONTAINING
An inquiry into the Nature, Cauſes, and Cure, of that Diſeaſe.
Together with
A Critical and Chronological View of what has been publiſhed on the ſubject.
By JAMES LIND, M. D.
Fellow of the Royal College of Phyſicians in Edinburgh.
EDINBURGH:
Printed by SANDS, MURRAY, and COCHRAN.
For A. KINCAID & A. DONALDSON.
MDCCLIII.

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Image: Wellcome Collection, via Wikimedia Commons. Lind J. A Treatise of the Scurvy. 1753.

CONTEXT

Where it fails

The coach gives the new boots to his best players
and they score more goals

Therefore: the boots work?

Illustrative example.

CONTEXT

Limitation of a non-randomised controlled trial

Non-random allocation

Where it failed

The coach gave new boots to the strongest players. Attributing their extra goals to the boots ignores the advantage those players already had.

Selection bias

The groups differed before the treatment began, so the comparison was never fair.

Illustrative example.

CONTEXT

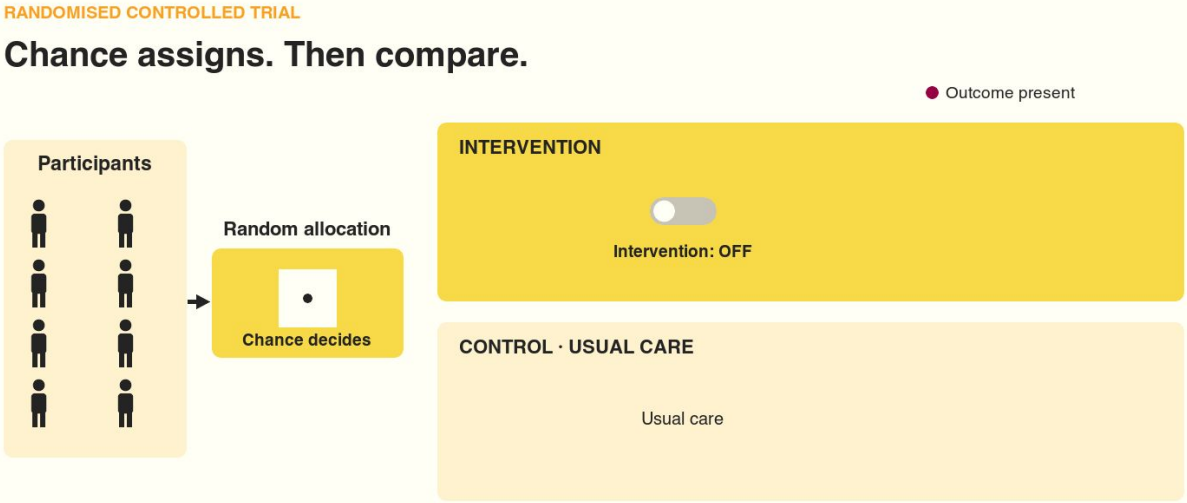
2 Intervene

Assign the treatment on purpose, then compare.

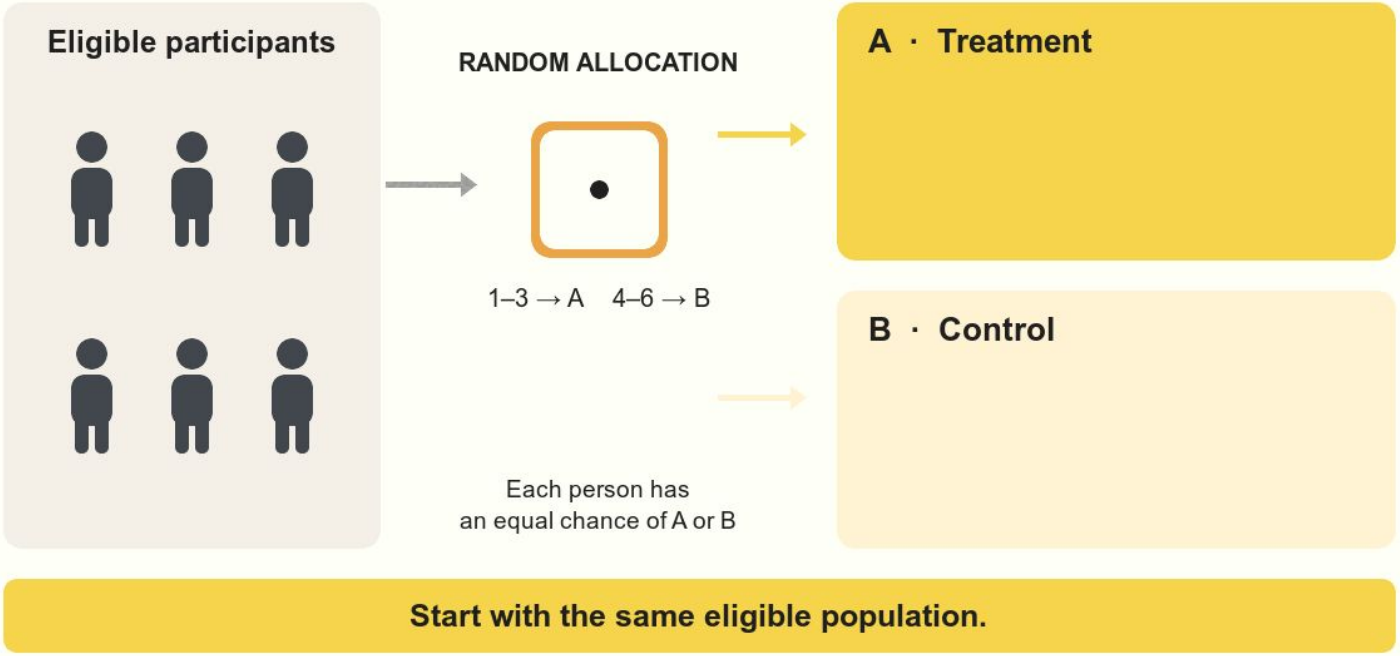
2.1 Before-and-after study

2.2 Controlled trial

2.3 Randomised trial



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.


Eligible participants
RANDOM ALLOCATION
1–3 → A
4–6 → B
A · Treatment
B · Control
Each person has
an equal chance of A or B
Start with the same eligible population.

CONTEXT

Example: the polio vaccine

UNITED STATES, 1954

400,000 children get vaccine or placebo, by chance

paralytic polio: 33 of about 200,000 vaccinated, 115 of about 200,000 on placebo

Therefore: the vaccine works



Image: iron lung ward, Rancho Los Amigos Hospital, via Wikimedia Commons. Francis T et al. Evaluation of the 1954 poliomyelitis vaccine trials. Am J Public Health 1955;45(5 Pt 2):1–63.

CONTEXT

2003: nobody had tested the parachute

In 2003, no randomised trial had ever tested the parachute
nobody would agree to be in the control group

Therefore: **some questions cannot be randomised**

Smith GCS, Pell JP. Parachute use to prevent death and major trauma. BMJ 2003;327:1459–1461.

CONTEXT

Limitation of a randomised trial

Ethical limits on randomisation

Where it failed

To test parachutes in a high-altitude jump, some participants would have to jump without one. That danger rules out the comparison, even though the question is easy to formulate.

Smith GCS, Pell JP. Parachute use to prevent death and major trauma. BMJ 2003;327:1459–1461.

CONTEXT

But actually, some people did



RESEARCH

Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial

Robert W Yeh,[1] Linda R Valsdottir,[1] Michael W Yeh,[2] Changyu Shen,[1] Daniel B Kramer,[1] Jordan B Strom,[1] Eric A Secemsky,[1] Joanne L Healy,[1] Robert M Domeier,[3] Dhruv S Kazi,[1] Brahmajee K Nallamothu[4] On behalf of the PARACHUTE Investigators

Peer reviewed and published in the BMJ Christmas issue, 2018.

Yeh RW et al. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. BMJ 2018;363:k5094 (Christmas issue).

CONTEXT

Who signed up

Table 1 | Baseline characteristics of participants randomized to parachute versus control. Values are numbers (percentages) unless stated otherwise

Characteristics	Parachute	Control
Total	11 (100)	12 (100)
Demographics		
Median (SD) age (years)	38.1 (8.7)	38.6 (11.0)
Women	4 (36)	6 (50)
Men	7 (64)	6 (50)
Ethnic group:		
American Indian or Alaska Native	0 (0)	0 (0)
East Asian or South Asian	4 (36)	4 (33)
Black or African American	0 (0)	0 (0)
More than one race	0 (0)	0 (0)
White	7 (64)	8 (67)
Mean (SD) height (cm)	171.8 (9.1)	171.7 (8.4)
Mean (SD) weight (kg)	75.9 (24.4)	74.6 (13.0)
Medical history		
Broken bones	4 (36)	5 (42)
Acrophobia	3 (27)	6 (50)
Parachute use	3 (27)	0 (0)
Family history of parachute use	2 (18)	0 (0)
Frequent flier (average >4 flights per month)	0 (0)	4 (33)
Flight		
International v domestic:		
International	0 (0)	0 (0)
Domestic	11 (100)	12 (100)
Aircraft type:		
Jetliner	0 (0)	0 (0)
Biplane	5 (46)	6 (50)
Helicopter	6 (55)	6 (50)
Mean (SD) velocity (km/h)	0 (0)	0 (0)
Mean (SD) altitude (m)	0.6 (0.1)	0.6 (0.1)

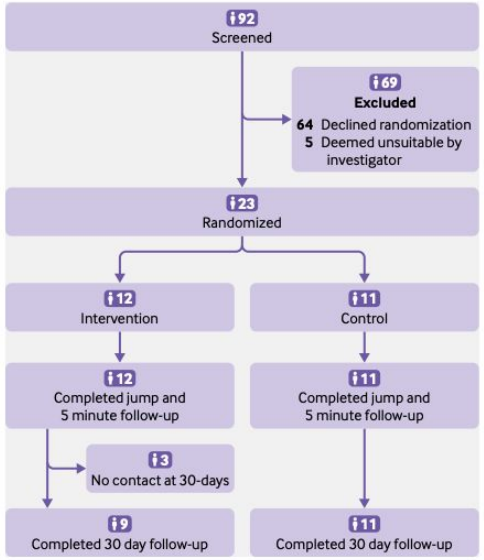


Fig 1 | Study flow diagram

92 people asked

69 said no

23 jumped

Mean altitude: 0.6 m

Mean speed: 0 km/h

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Yeh RW et al. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. BMJ 2018;363:k5094 (Christmas issue).

CONTEXT

What they found

WHAT IS ALREADY KNOWN ON THIS TOPIC

Parachutes are routinely used to prevent death or major traumatic injury among individuals jumping from aircraft, but their efficacy is based primarily on biological plausibility and expert opinion

No randomized controlled trials of parachute use have yet been attempted, presumably owing to a lack of equipoise

WHAT THIS STUDY ADDS

This randomized trial of parachute use found no reduction in death or major injury compared with individuals jumping from aircraft with an empty backpack

Lack of enrolment of individuals at high risk could have influenced the results of the trial



Fig 2 | Representative study participant jumping from aircraft with an empty backpack. This individual did not incur death or major injury upon impact with the ground

Therefore: randomised, but does it answer the question we care about?

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

Yeh RW et al. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. BMJ 2018;363:k5094 (Christmas issue).

CONTEXT

Control, placebo, sham

Control

The comparison strategy. It can be no treatment, a placebo or an active treatment.

Placebo

A dummy treatment that looks identical. It supports blinding but does not guarantee it.

Sham

A placebo procedure. In a knee trial, the tested surgery showed no advantage over skin cuts alone.

Moseley JB et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med 2002;347:81–88.

CONTEXT

ENGINEERING

Evaluate the whole clinical intervention



Clinical benefit → an intervention-level evaluation target. Predictive accuracy measures a different quantity.

Manz et al. 2023; CONSORT-AI intervention reporting.

CONTEXT

Vocabulary: what each term helps us decide

Causal question

✓ Association ✓ Confounding
✓ Counterfactual ✓ Treatment effect

Fair comparison

✓ Randomisation ✓ Comparator
✓ Placebo ✓ Sham

Study design

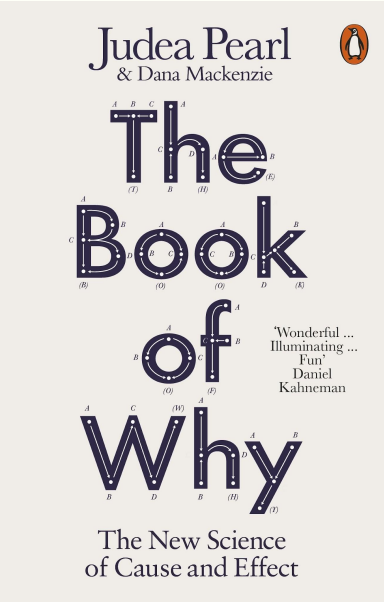
✓ Case series ✓ Cross-sectional
✓ Ecological ✓ Case-control
✓ Cohort ✓ Controlled trial ✓ RCT

Next week

Systematic review, meta-analysis
Equipoise, consent
Ethics review, monitoring

CONTEXT

Must read: The Book of Why



What additional knowledge lets us move from predicting an outcome to predicting the consequences of an action?

Why it belongs: prediction is not intervention.

Pearl & Mackenzie. The Book of Why, 2018. Actual cover: Penguin.

CONTEXT

ENGINEERING

From clinical questions to causal inference

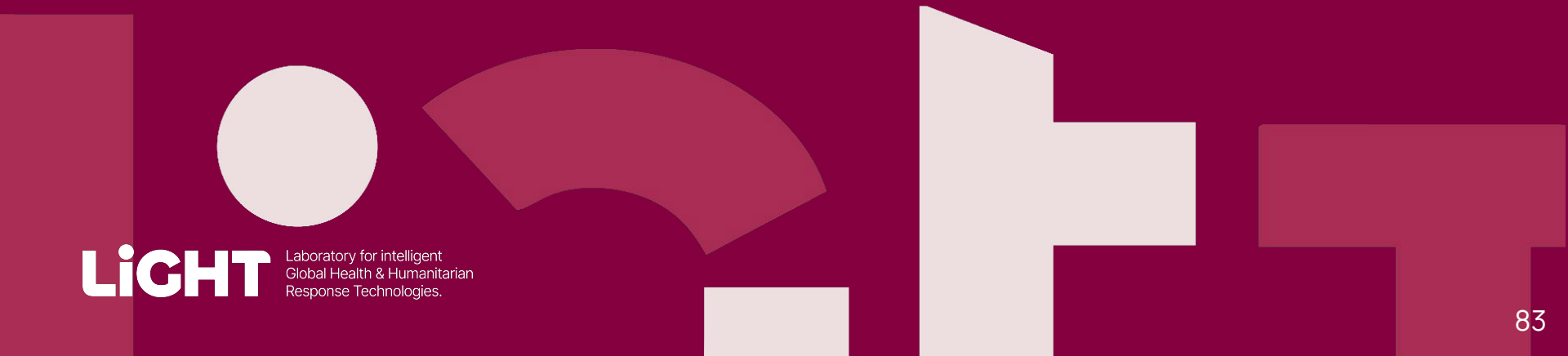
When does predicting what happens let us predict what would happen if we changed care?

Questions?

EPFL

Causal inference

Lecturer: *Dr. Lars Klein*



LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

ENGINEERING

Outline

01 We built an intervention

02 A clinical trial

03 Sometimes you win the lottery

04 Sometimes you don't

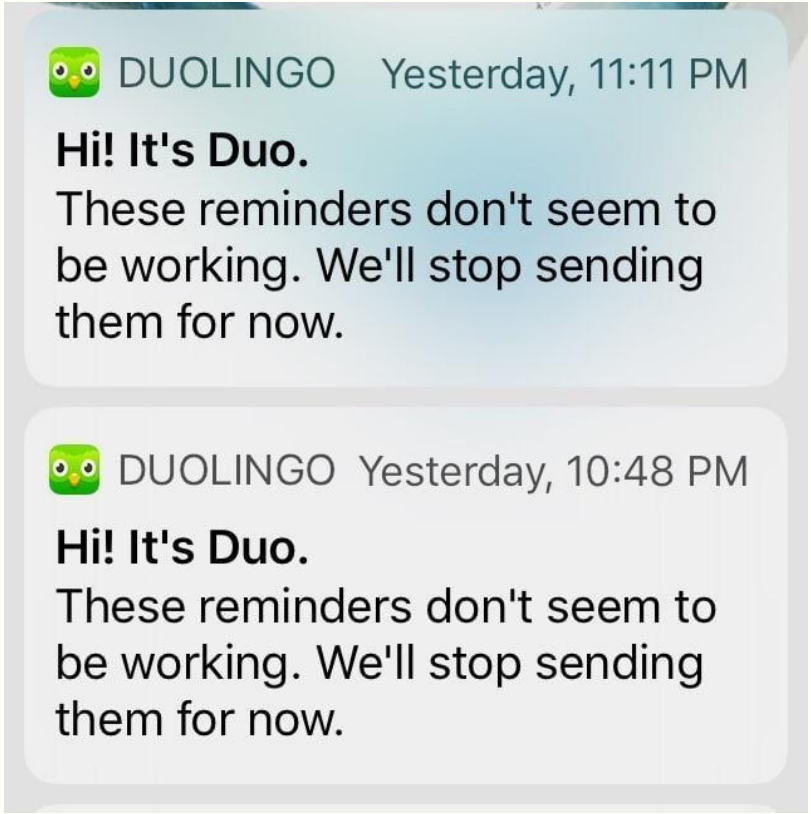
We built an intervention

How are you doing this week?

I’m having trouble taking my medication.

Can we call you to help?

Does it work?


DUOLINGO Yesterday, 11:11 PM
Hi! It's Duo.
These reminders don't seem to be working. We'll stop sending them for now.
DUOLINGO Yesterday, 10:48 PM
Hi! It's Duo.
These reminders don't seem to be working. We'll stop sending them for now.

WelTel Kenya1

Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial

Richard T Lester, Paul Ritvo, Edward J Mills, Antony Kariri, Sarah Karanja, Michael H Chung, William Jack, James Habyarimana, Mohsen Sadatsafavi, Mehdi Najafzadeh, Carlo A Marra, Benson Estambale, Elizabeth Ngugi, T Blake Ball, Lehana Thabane, Lawrence J Gelmon, Joshua Kimani, Marta Ackers, Francis A Plummer

The Lancet, 2010

What did they compare?

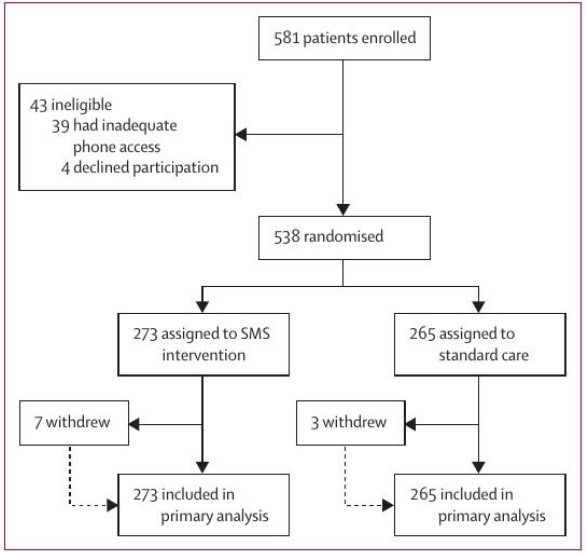


Figure 1: **Trial profile**

Weekly SMS support
+ usual care

versus

Usual care

Did it help?

	SMS group (number [%])	Control group (number [%])	RR (95% CI)*	p value
Primary outcome				
Intention-to-treat analysis†				
Self-reported adherence (>95%)	168 (62%)	132 (50%)	0·81 (0·69–0·94)	0·006
Viral suppression (<400 copies per mL)	156 (57%)	128 (48%)	0·85 (0·72–0·99)	0·04

RR refers to non-adherence or virologic failure.

Is randomization ethical?

Equipoise

Genuine uncertainty about which option is better.
Both groups received HIV treatment.

Is randomization ethical?

Equipoise

Genuine uncertainty about which option is better.
Both groups received HIV treatment.

Often, you simply can’t randomize!

Sometimes you win the lottery



Sometimes you ~~win the lottery~~ find a natural experiment



The Japanese console lottery

nature human behaviour

Article

https://doi.org/10.1038/s41562-024-01948-y

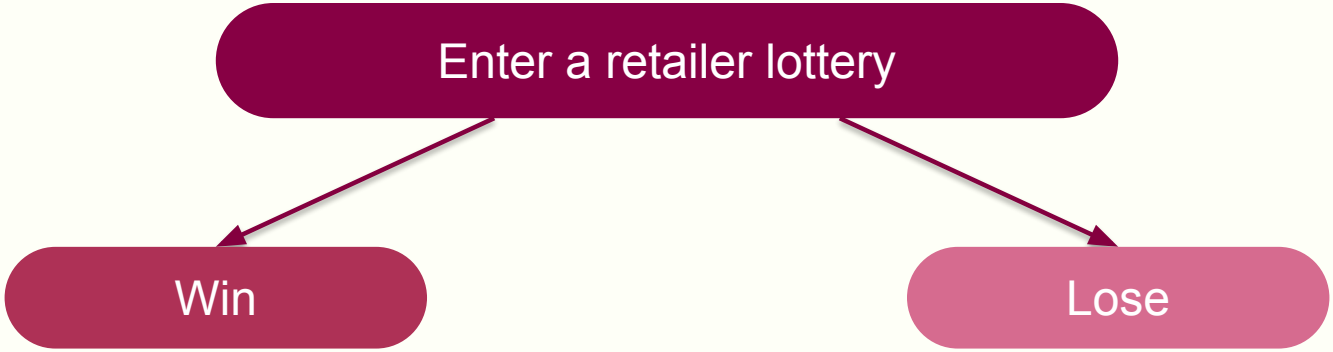
Causal effect of video gaming on mental well-being in Japan 2020–2022

Received: 22 June 2023

Accepted: 5 July 2024

Hiroyuki Egami [1,2] ✉, Md. Shafiur Rahman [3,4], Tsuyoshi Yamamoto [5], Chihiro Egami [6] & Takahisa Wakabayashi [7]

What did the lottery decide?



Opportunity to buy a console

No purchase opportunity from that lottery

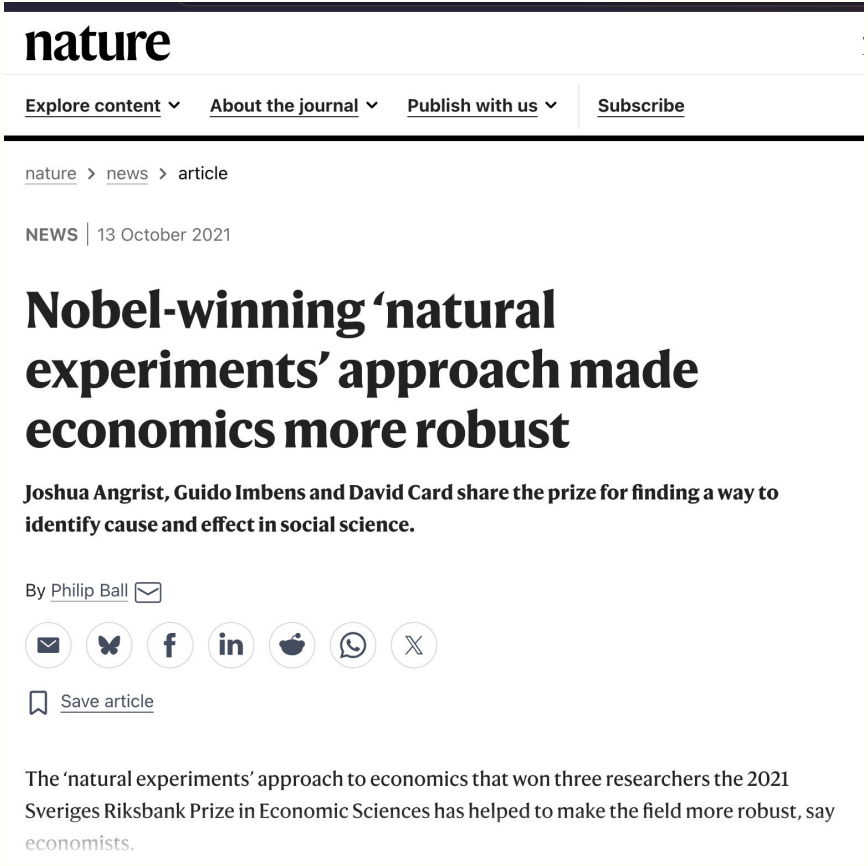
What happened to well-being?

Switch ownership
0.60 SD less psychological distress (95% CI 0.38–0.82)

PS5 ownership
0.12 SD less psychological distress (95% CI 0.03–0.21)

PS5 ownership
0.23 SD greater life satisfaction (95% CI 0.16–0.31)

Estimated effects for people whose ownership changed because of the lottery.



nature

Explore content | About the journal | Publish with us | Subscribe

nature > news > article

NEWS | 13 October 2021

Nobel-winning 'natural experiments' approach made economics more robust

Joshua Angrist, Guido Imbens and David Card share the prize for finding a way to identify cause and effect in social science.

By Philip Ball

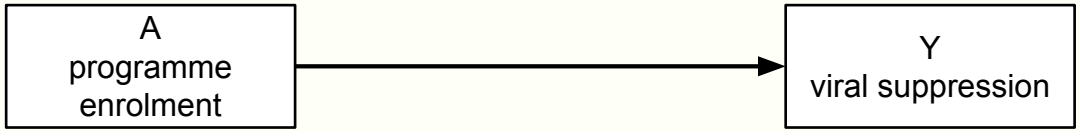
Save article

The 'natural experiments' approach to economics that won three researchers the 2021 Sveriges Riksbank Prize in Economic Sciences has helped to make the field more robust, say economists.

Sometimes you don't win the lottery



Does our HIV support programme help?



$A \rightarrow Y$

Patients choose whether to enrol.

Everyone continues usual HIV care.

Average treatment effect

$ATE = E[Y^1] - E[Y^0]$

Y^1: outcome under enrolment

Y^0: outcome without enrolment

We observe only one outcome for each patient.

What could make this comparison difficult?

Observation versus intervention

$P(Y = 1 \mid A = 1)$

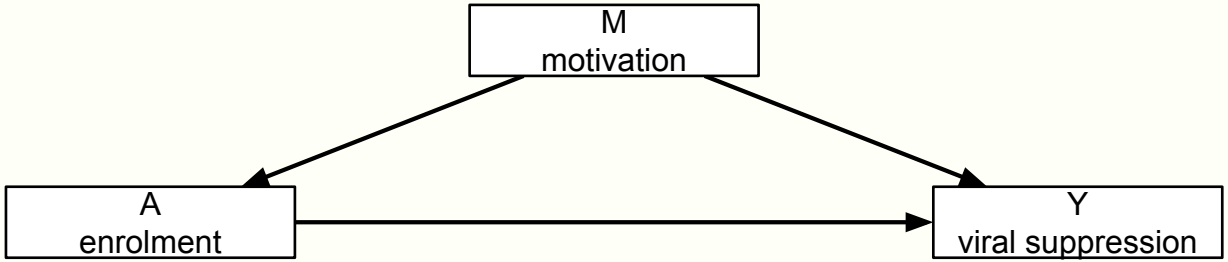
Suppression among patients who actually enrolled.

$P(Y = 1 \mid \text{do}(A = 1))$

Suppression if everyone were assigned to enrol.

For binary Y: $E[Y^1] = P(Y = 1 \mid \text{do}(A = 1))$

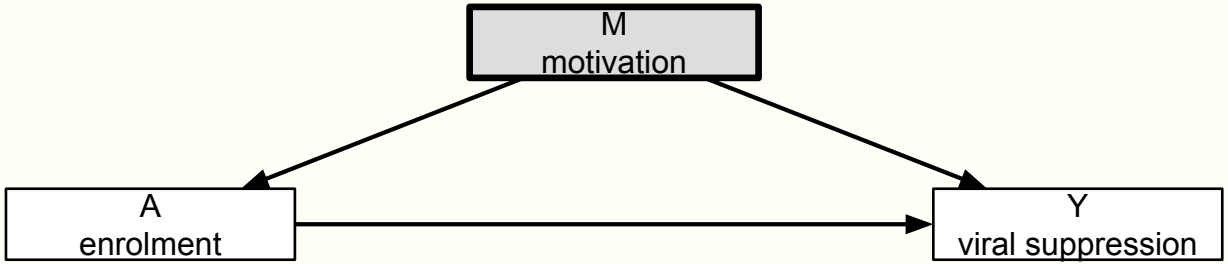
Confounder



$A \rightarrow Y$ $A \leftarrow M \rightarrow Y$

What should we do?

Conditioning on motivation



Compare patients with the same baseline motivation.
A ← M → Y is blocked.

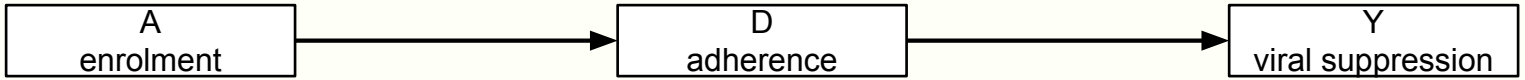
Average treatment effect, adjusted for motivation

ATE = $\sum_m$ [E[Y | A=1, M=m] − E[Y | A=0, M=m]] P(M=m)

Compare within each motivation group.

Average using the population's distribution of motivation.

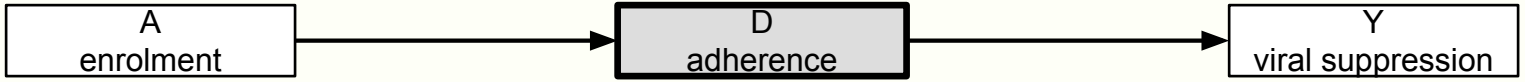
Mediator



$A \rightarrow D \rightarrow Y$

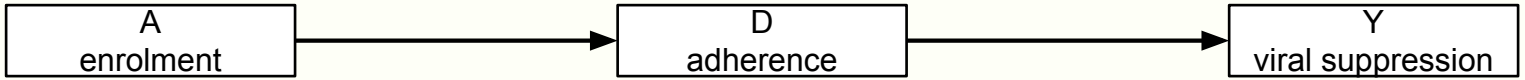
Do we need to adjust for anything?

Condition on adherence?



Compare patients with the same adherence?

No adjustment for adherence

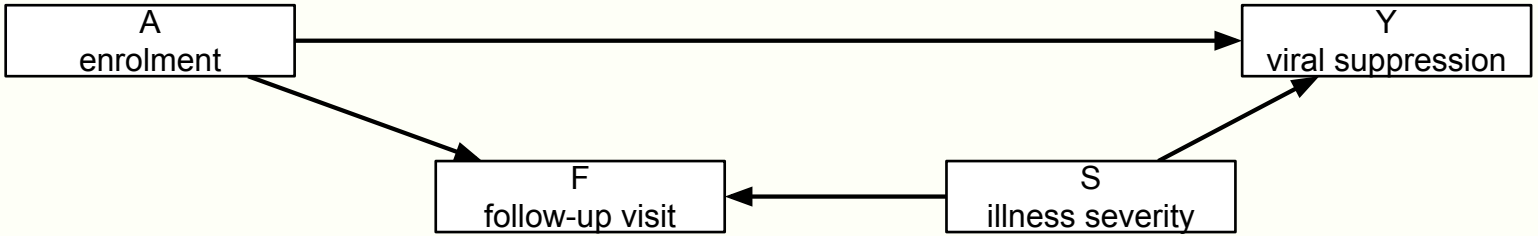


ATE = $E[Y^1] - E[Y^0]$

Conditioning on adherence blocks the mechanism we care about.

For the total effect, preserve this pathway.

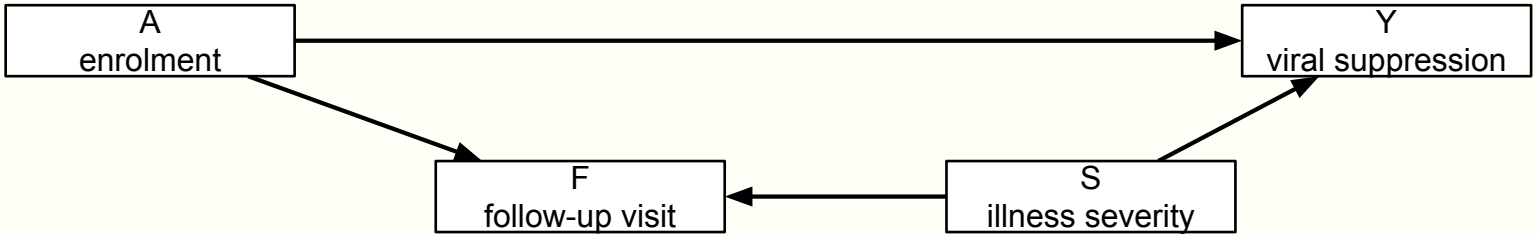
Collider



$A \rightarrow Y$ $A \rightarrow F \leftarrow S \rightarrow Y$

Do we condition on anything here?

Collider



$A \rightarrow Y$ $A \rightarrow F \leftarrow S \rightarrow Y$

No additional adjustment is needed in this graph.

Which variables should we condition on?

There may be several variables to consider.

Z: the collection of variables we condition on.

How do we choose them?

Z can contain several variables

Adjustment set: {phone access P, motivation M}

Z = (P, M) **z = (p, m)**

$\sum_z$ **f(z) P(z) =** $\sum_p \sum_m$ **f(p, m) P(p, m)**

Sum over joint values, not over variables.

If both are binary: (0,0), (0,1), (1,0), (1,1)

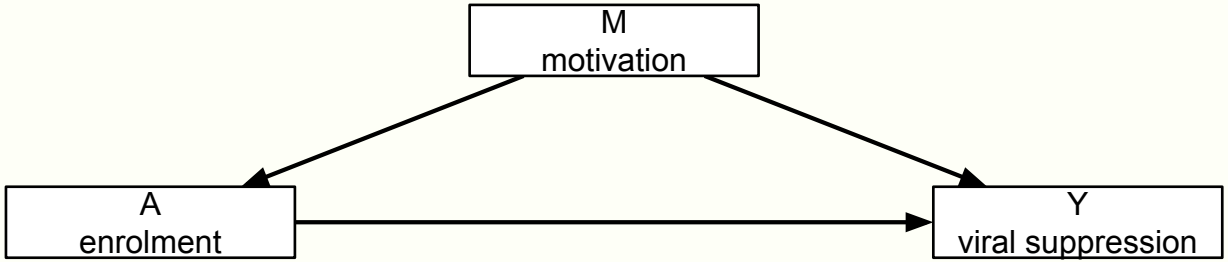
The adjustment formula

$P(Y=1 \mid do(A=a)) = \sum_z P(Y=1 \mid A=a, Z=z)\, P(Z=z)$

A: enrolment variable. a: a chosen value, 0 or 1.

For a suitable adjustment set Z.

What is a backdoor path?



Causal path: A → Y
Backdoor path: A ← M → Y

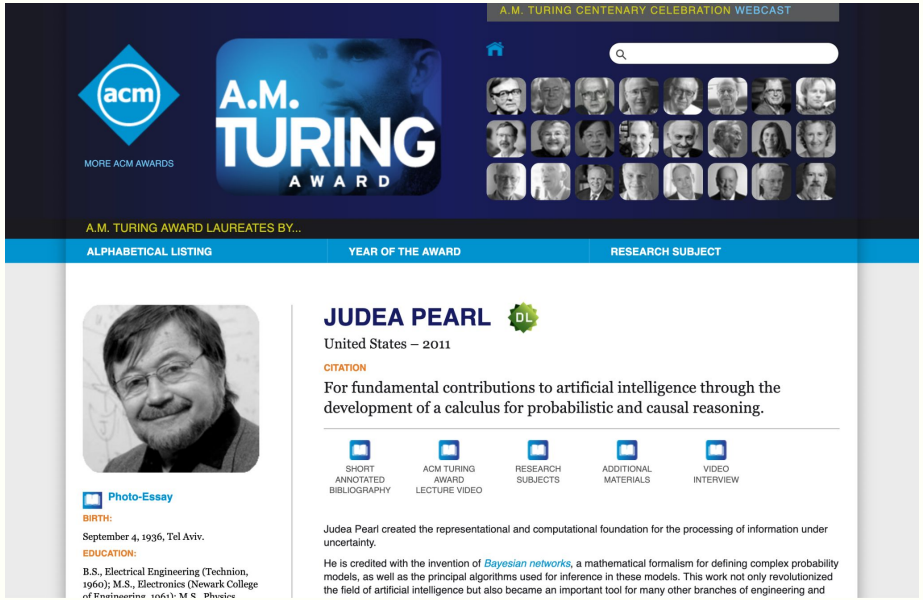
The backdoor criterion

A sufficient rule for choosing an adjustment set Z.

1. No variable in Z is a descendant of A.
2. Conditioning on Z blocks every path from A to Y that starts with an arrow into A.

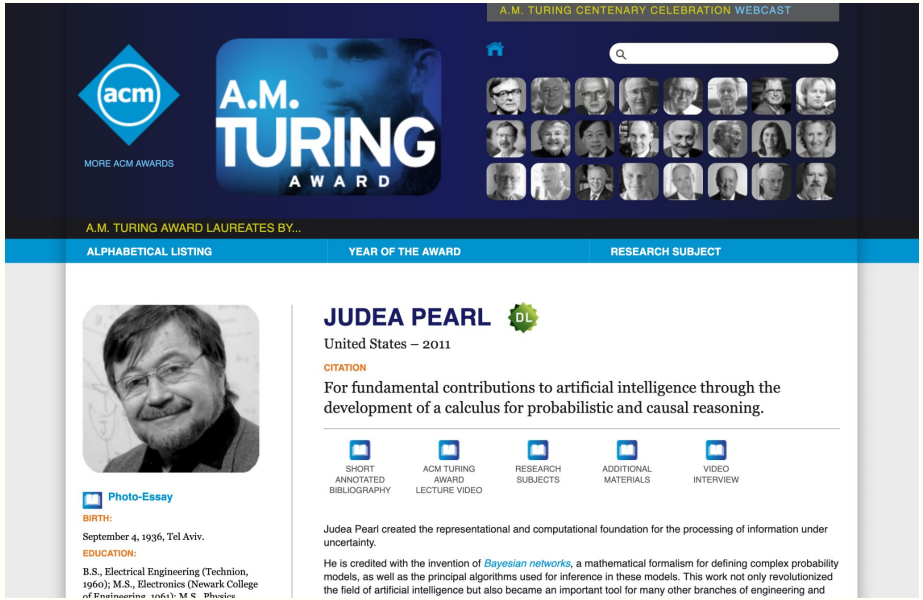
Judea Pearl, Causality, Definition 3.3.1 and Theorem 3.3.2

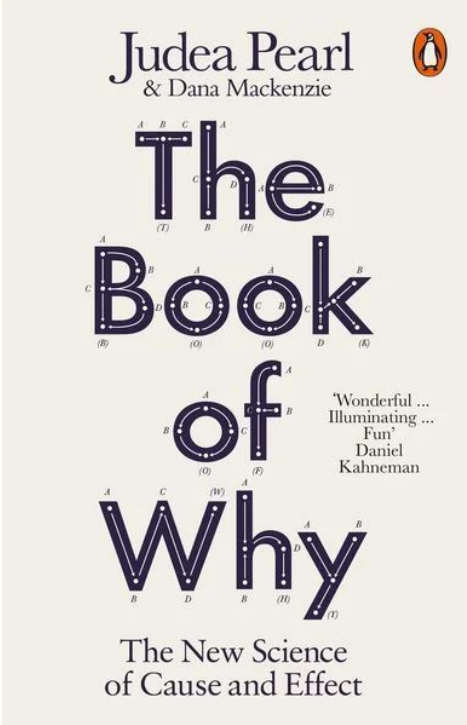
Judea Pearl


A.M. TURING CENTENARY CELEBRATION WEBCAST
acm
MORE ACM AWARDS
A.M. TURING AWARD
A.M. TURING AWARD LAUREATES BY...
ALPHABETICAL LISTING
YEAR OF THE AWARD
RESEARCH SUBJECT
JUDEA PEARL
United States – 2011
CITATION
For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.
SHORT ANNOTATED BIBLIOGRAPHY
ACM TURING AWARD LECTURE VIDEO
RESEARCH SUBJECTS
ADDITIONAL MATERIALS
VIDEO INTERVIEW
Photo-Essay
BIRTH:
September 4, 1936, Tel Aviv.
EDUCATION:
B.S., Electrical Engineering (Technion, 1960); M.S., Electronics (Newark College
Judea Pearl created the representational and computational foundation for the processing of information under uncertainty.
He is credited with the invention of Bayesian networks, a mathematical formalism for defining complex probability models, as well as the principal algorithms used for inference in these models. This work not only revolutionized the field of artificial intelligence but also became an important tool for many other branches of engineering and

ENGINEERING

Judea Pearl





LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.

When is a path blocked?

Non-collider: → **B** → **or** ← **B** →

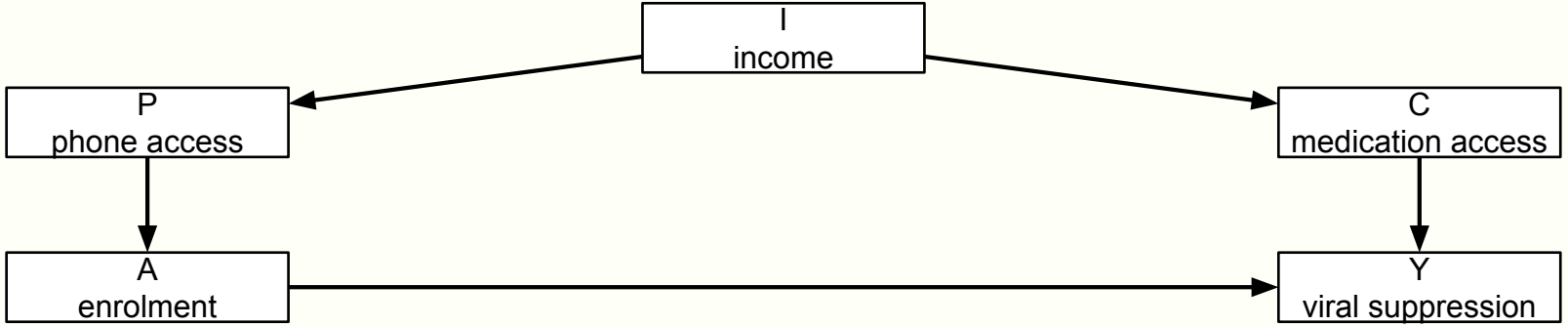
Conditioning on B blocks the path.

Collider: → **K** ←

The path is blocked if neither K,
nor any of its descendants is conditioned on.

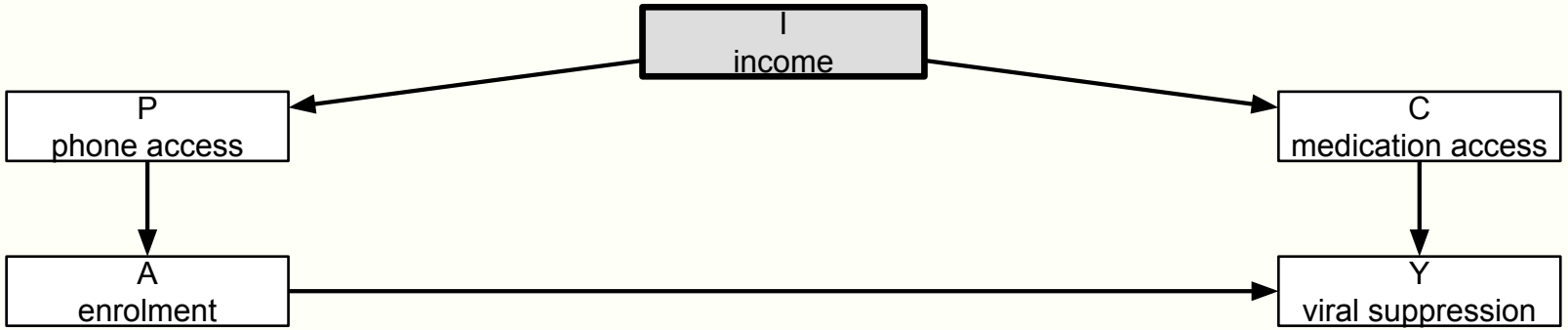
One blocking point is enough.

A longer backdoor path



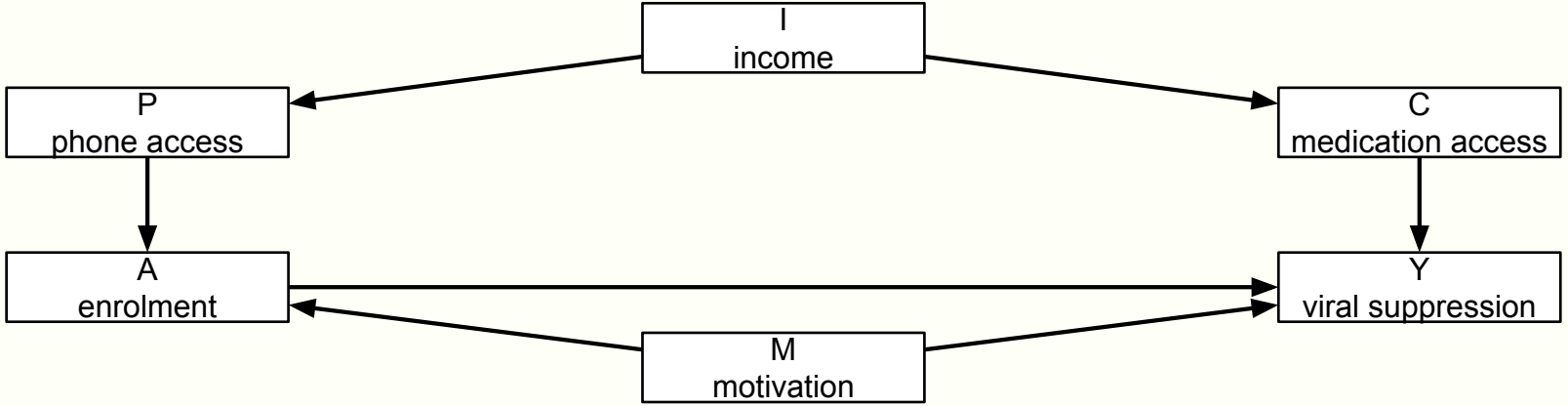
$A \leftarrow P \leftarrow I \rightarrow C \rightarrow Y$

Where can we block it?



Adjust for phone access, income, or medication access.
Grey = conditioned on

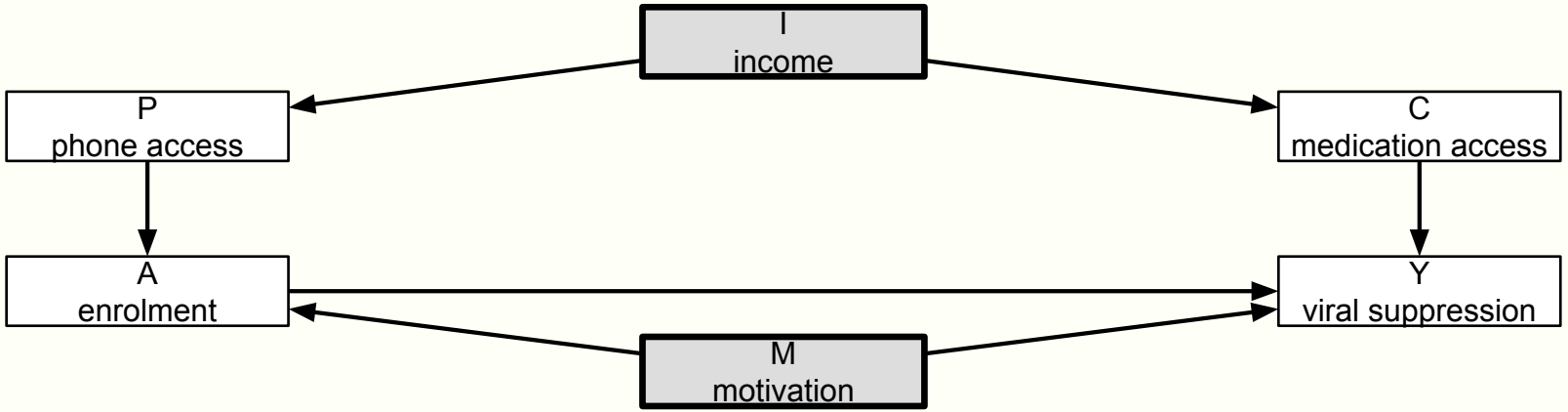
A second backdoor path



$A \leftarrow P \leftarrow I \rightarrow C \rightarrow Y$

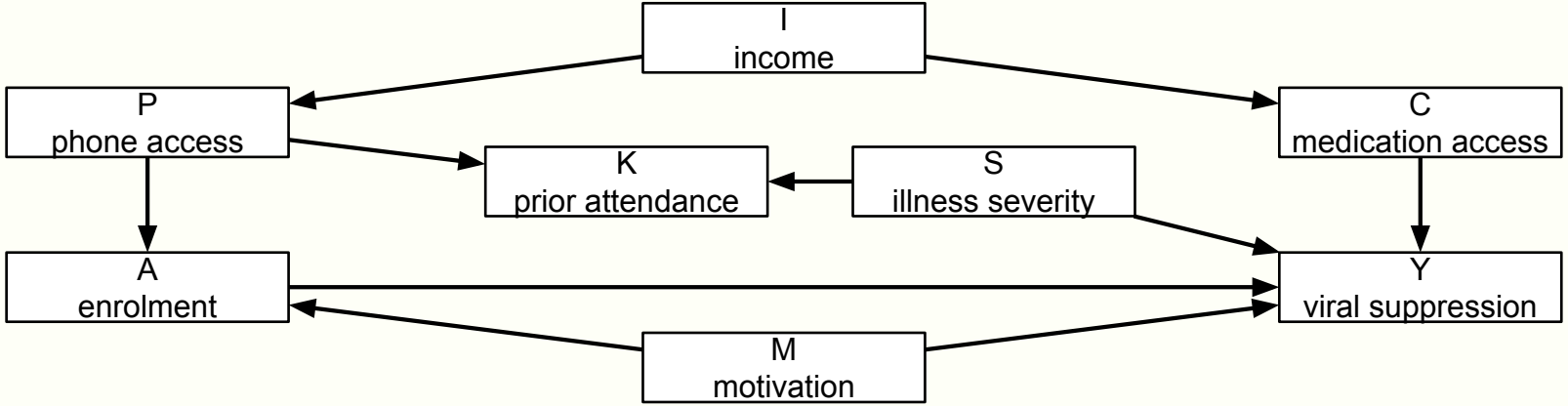
$A \leftarrow M \rightarrow Y$

Block both paths



Adjustment set: {income, motivation}

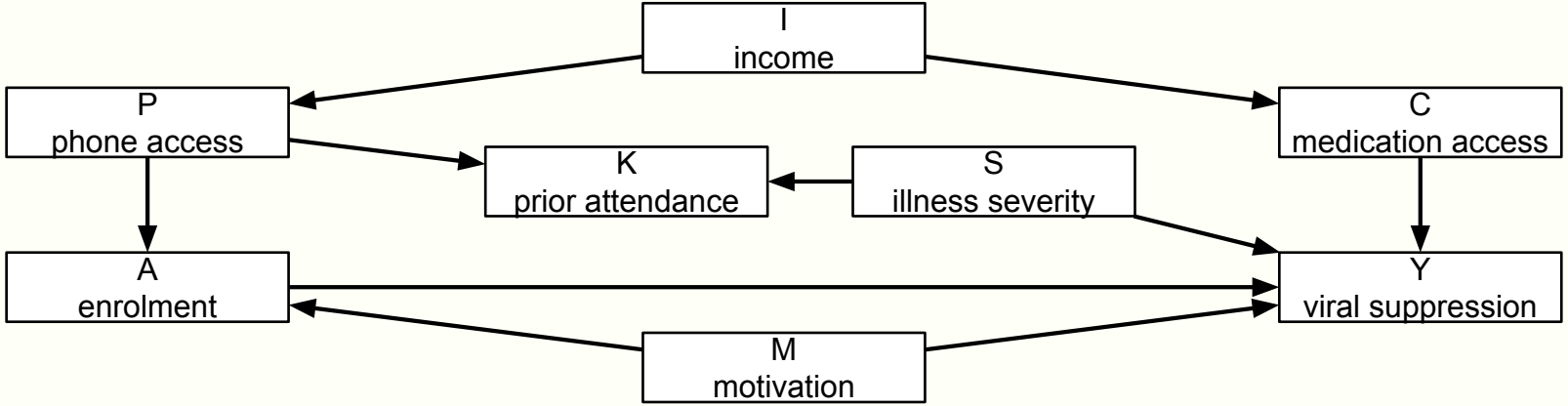
Condition on nothing



$A \leftarrow P \rightarrow K \leftarrow S \rightarrow Y$

Is this path blocked?

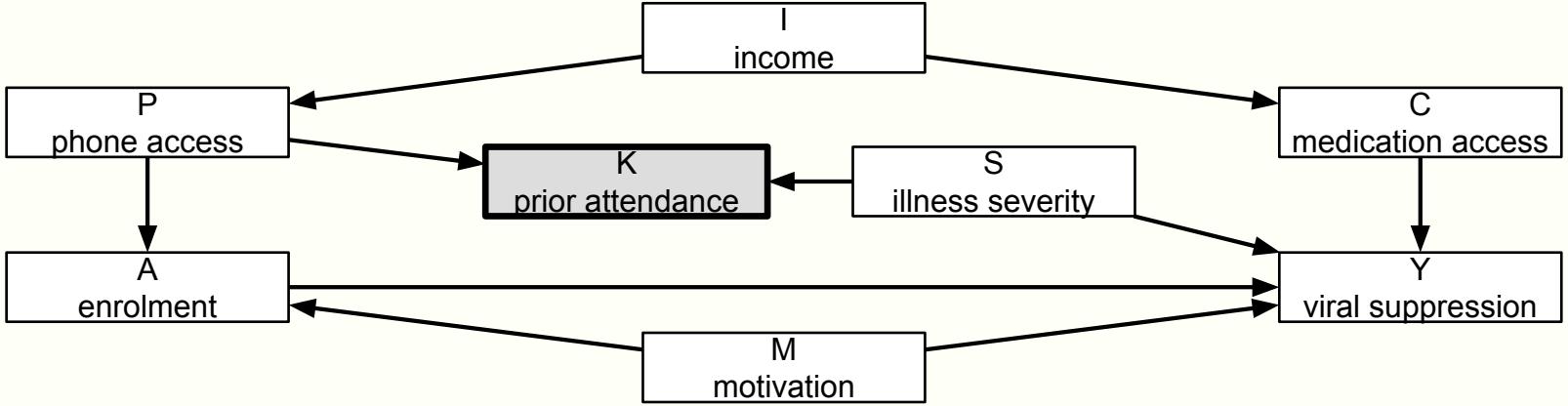
Condition on nothing



$A \leftarrow P \rightarrow K \leftarrow S \rightarrow Y$

Yes. Blocked at K.

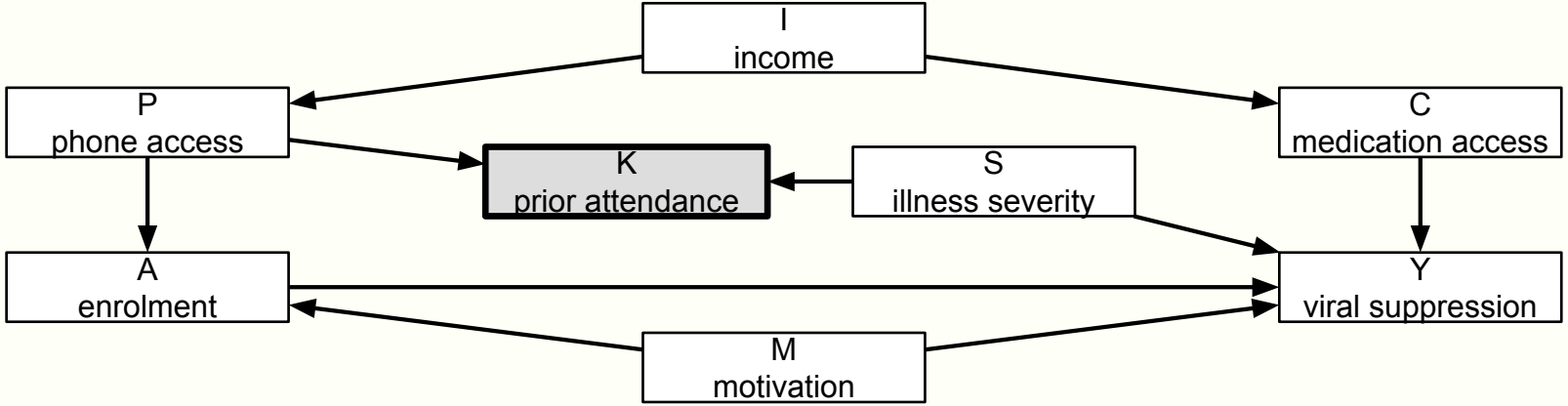
Condition on K



$A \leftarrow P \rightarrow K \leftarrow S \rightarrow Y$

Is this path blocked?

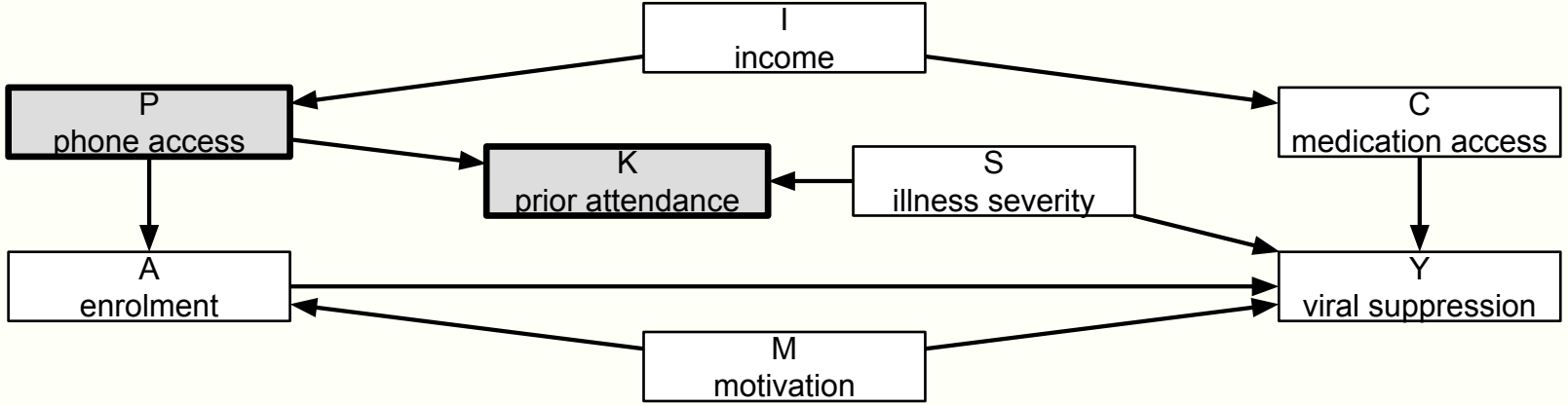
Condition on K



$A \leftarrow P \rightarrow K \leftarrow S \rightarrow Y$

No. The path is open.

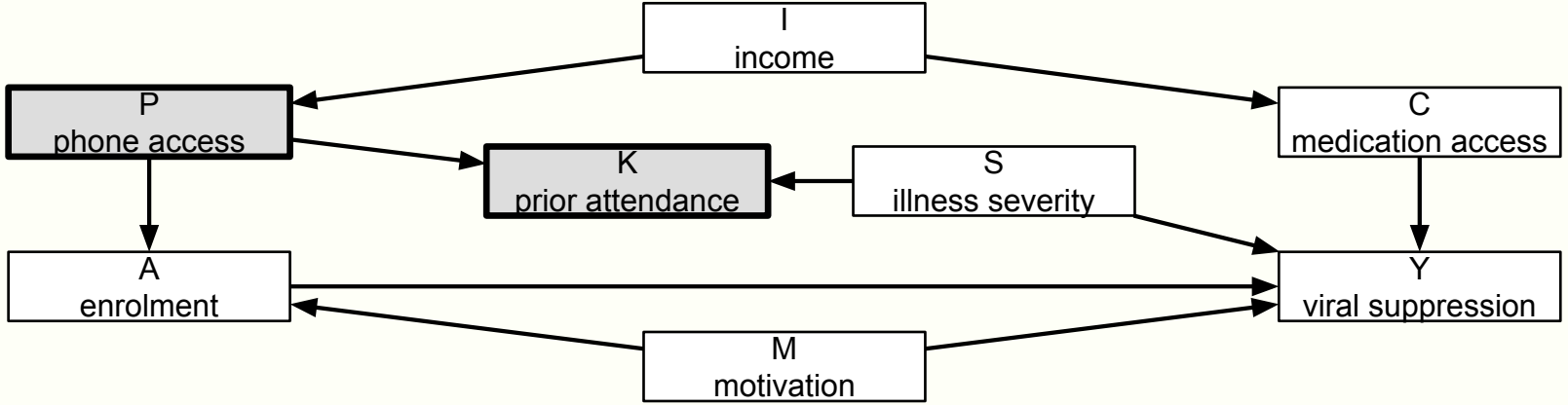
Condition on P and K



$A \leftarrow P \rightarrow K \leftarrow S \rightarrow Y$

Is this path blocked?

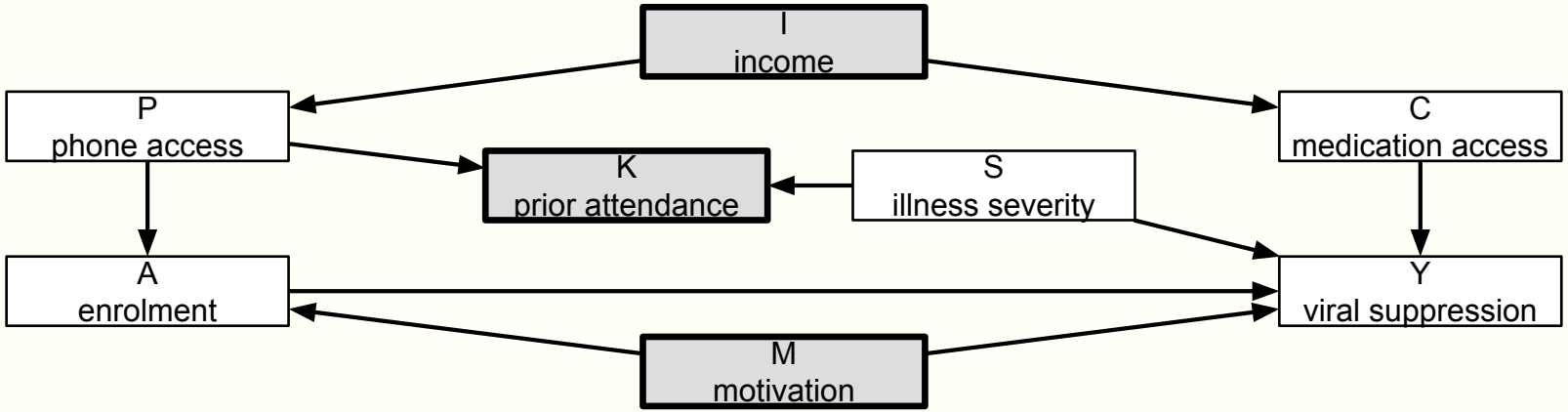
Condition on P and K



A ← P → K ← S → Y

Yes. Blocked at P.

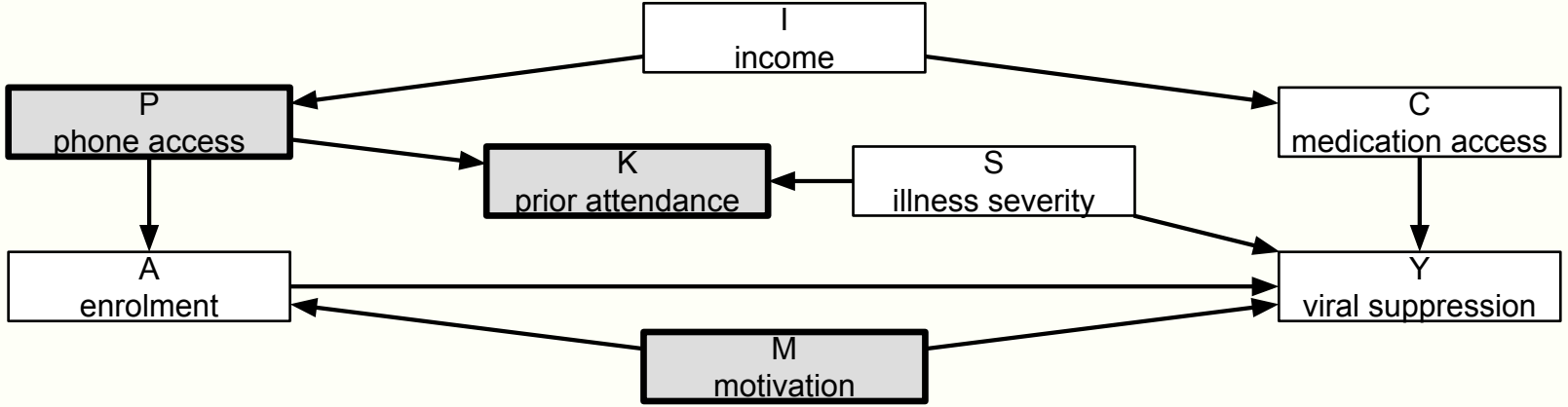
Conditioning on the collider



{income, motivation, prior attendance}

The path A ← P → K ← S → Y is now open.

Block the path somewhere else



Phone access blocks the route, even with K conditioned on.
{phone access, motivation} already suffices.

G-computation

$P(Y=1 \mid do(A=a)) = \sum_z P(Y=1 \mid A=a, Z=z)\, P(Z=z)$

Estimate the outcome probabilities, then average over the population.

Inputs: enrolment, phone access, motivation.

Output: probability of viral suppression.

The outcome model and its fitted estimate

$\mu(a, z) := P(Y = 1 \mid A = a, Z = z)$

The conditional probability we want to estimate.

$\hat{\mu}(a, z)$**: our fitted estimate**

Fit using observed enrolment, characteristics and outcomes.

The hat means “estimated”.

Predict both scenarios for every patient

Patient	Predicted suppression under enrolment	Predicted suppression without enrolment
1	$\hat{\mu}(1, Z_1)$	$\hat{\mu}(0, Z_1)$
2	$\hat{\mu}(1, Z_2)$	$\hat{\mu}(0, Z_2)$
…	…	…

Patient i: $Z_i = (P_i, M_i)$

Same patients and baseline values. Change only enrolment.

G-computation: average and subtract

Average the enrolment column.

Subtract the average of the no-enrolment column.

Estimated ATE = $\frac{1}{n}\sum_{i=1}^{n}[\hat{\mu}(1, Z_i) - \hat{\mu}(0, Z_i)]$

n = number of patients

These averages estimate the outcomes under the two interventions.

Where did P(z) go?

Profiles: $\sum_z \mu(a, z) P(Z = z)$

Patients: $\frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(a, Z_i)$

A profile present in 60% of patients appears in 60% of rows.

It contributes 60% of the average.

The sample supplies our estimate of P(z).

Did someone say *Propensity Scores* ?

So far, we've modeled the probability of an outcome.

Alternatively, you can model the probability of treatment.

A very common practice is matching up treated/untreated participants based on their propensity score.

100 patients

Motivation	Enrolled: successful / total	Not enrolled: successful / total
High	54 / 60	16 / 20
Low	5 / 10	4 / 10
Total	59 / 70	20 / 30

We have 100 patients: 80 with high motivation and 20 with low motivation. “Successful” means virally suppressed.

A naive comparison…

Just compare the observed success rates.

Enrolled: 59/70 = 84.3%

Not enrolled: 20/30 = 66.7%

Difference: 17.6 percentage points.

The groups have different compositions

Motivation	Enrolled: successful / total	Not enrolled: successful / total
High	54 / 60	16 / 20
Low	5 / 10	4 / 10
Total	59 / 70	20 / 30

Highly motivated patients constitute 86% of the enrolled group and only 67% of the non-enrolled group.

Calculate the propensity scores

Within the high-motivation group, 60 of 80 enrol: propensity score 0.75.

Within the low-motivation group, 10 of 20 enrol: propensity score 0.5.

$e(z) = P(A=1 \mid Z=z)$

Weight the patients

Group	Weight per patient	Weighted patients	Weighted successes
Enrolled, high	1/0.75 = 4/3	60 × 4/3 = 80	54 × 4/3 = 72
Enrolled, low	1/0.5 = 2	10 × 2 = 20	5 × 2 = 10
Not enrolled, high	1/0.25 = 4	20 × 4 = 80	16 × 4 = 64
Not enrolled, low	1/0.5 = 2	10 × 2 = 20	4 × 2 = 8

Use the inverse probability of the treatment each patient actually received.

Now the compositions match

Group	Weight per patient	Weighted patients	Weighted successes
Enrolled, high	1/0.75 = 4/3	60 × 4/3 = 80	54 × 4/3 = 72
Enrolled, low	1/0.5 = 2	10 × 2 = 20	5 × 2 = 10
Not enrolled, high	1/0.25 = 4	20 × 4 = 80	16 × 4 = 64
Not enrolled, low	1/0.5 = 2	10 × 2 = 20	4 × 2 = 8

Both weighted groups now represent the population's 80 high / 20 low motivation composition.

These are weighted totals, not additional patients.

Compare the weighted success rates

Enrolled: (72+10)/(80+20) = 82%

Not enrolled: (64+8)/(80+20) = 72%

Difference: 10 percentage points.

Before and after adjustment

The estimate falls from 17.6 to 10 percentage points after adjusting for motivation.

Interpreting that adjusted difference causally assumes that motivation is sufficient to control confounding in this example!

The backdoor criterion still applies!

G-computation and weighting

G-computation models outcomes and averages predictions.

Weighting models enrolment and reweights observed outcomes.

Both need a justified **adjustment set**

Takeaway messages

Choose adjustment variables from the causal graph.

Good predictions do not establish a causal effect.

And we haven’t even talked about the models.
(The usual suspects are models that like tabular data. Logistic Regression, Random Forests, …)

Questions?

Building intelligent Global Health and Humanitarian Response Technologies

Thank You

Except where otherwise indicated, original content in this presentation is © 2026 Hartley, Keitel, Klein, Sasu, Suttels, Arni, Kesmen, Nemo and licensed under CC BY 4.0: https://creativecommons.org/licenses/by/4.0/.

Third-party materials are excluded from this license and remain subject to their indicated rights.

LiGHT Laboratory for intelligent Global Health & Humanitarian Response Technologies.