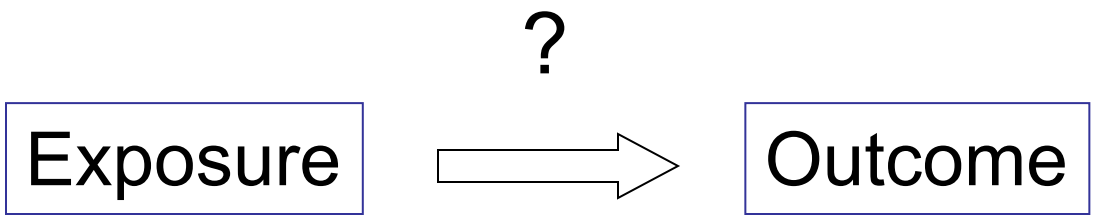
Causality & causal inference

Madhukar Pai, MD, PhD
McGill University
madhukar.pai@mcgill.ca

McGill

Of the 3 types of knowing ("gnosis") etio-gnosis (causality) is the central concern of epidemiology

- Most fundamental application of epidemiology: to identify etiologic (causal) associations between exposure(s) and outcome(s)



What is a cause?

- “Cause of a disease event is an event, condition or characteristic that preceded the disease event and without which the disease event either would not have occurred at all or would not have occurred until some other time.”
  - » [Rothman & Greenland, 1998]

Cause and effect?

Tuberculosis vaccine ‘potential game-changer’ in Covid-19 fight

Countries with high BCG vaccination rates have fewer coronavirus deaths, study shows

Coronavirus: Is TB vaccines a silver bullet in fight against COVID-19

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

CSH Cold Spring Harbor Laboratory BMJ Yale

HOME |

Search

Comments (145)

Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study

Aaron Miller, Mac Josh Reandelar, Kimberly Fasciglione, Violeta Roumenova, Yan Li, Gonzalo H Otazu

doi: https://doi.org/10.1101/2020.03.24.20042937

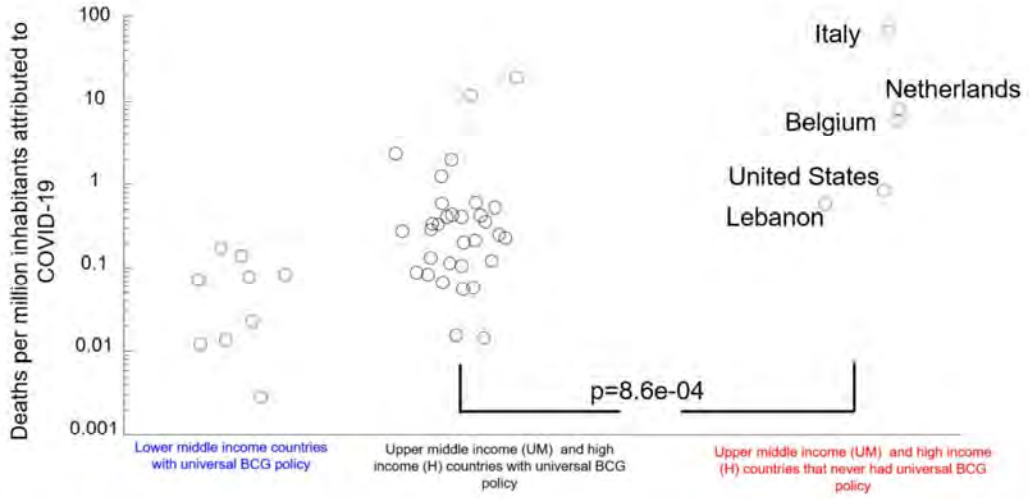


Figure 1: Higher death rates were presented in countries that never implemented a universal BCG vaccination policy.

Covid-19

Smokers seem less likely than non-smokers to fall ill with covid-19

That may point towards a way of treating it

Low incidence of daily active tobacco smoking in patients with symptomatic COVID-19 Preprint v3

Makoto Miyara[1], Florence Tubach[1], Valérie POURCHER[1], Capucine Morelot-Panzini[1], Julie Pernet[1], Julien Haroche[1], Said Lebbah[1], Elise Morawiec, Guy Gorochov[2], Eric Caumes[1], Pierre Hausfater[1], Alain COMBES[1], Thomas Similowski, Zahir Amoura[1]

"Our cross sectional study in both COVID-19 out- & inpatients strongly suggests that daily smokers have a very much lower probability of developing symptomatic or severe SARS-CoV-2 infection as compared to the general pop."

What is a cause?

- Cause
  - Must precede the effect (absolute requirement)
  - Can be either host, agent or environmental factors (e.g. characteristics, conditions, infection, actions of individuals, events, natural, social phenomena)
  - Can be either
    - positive = the presence of an exposure (e.g. radiation)
    - negative = the absence of exposure (e.g. vaccination)

- Should always be set up as a comparison:
  - "Cause is a category of a determinant, in relation to a particular reference category, capable of completing a sufficient cause in some instances in which the reference category is incapable of such completion" [OS Miettinen]

Compared to what?

- In an old movie, comedian Groucho Marx is asked: “Groucho, how’s your wife?”
- Groucho quips: “Compared to what?”

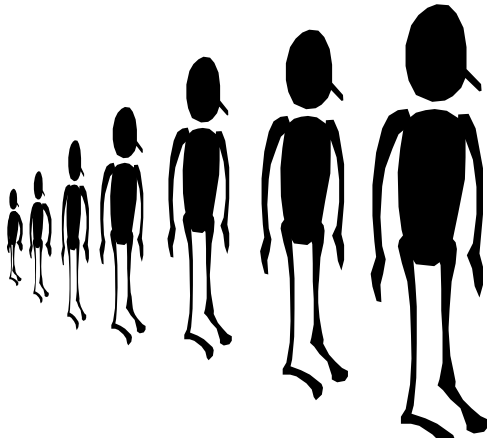
http://en.wikipedia.org

What is a causal effect?

- To determine a causal effect, we always need to set up a causal contrast (against some reference)
- Ideal "causal contrast" between exposed and unexposed groups:
  - "A causal contrast compares disease frequency under *two* exposure distributions, but in *one* target population during *one* etiologic time period"
  - If the ideal causal contrast is met, the observed effect is the "causal effect"

Maldonado & Greenland, Int J Epi 2002;31:422-29

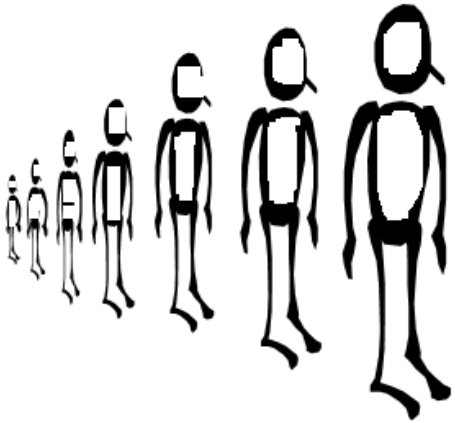
Ideal counterfactual comparison to determine causal effects



I_{exp}

Exposed cohort

"Initial conditions" are identical in the exposed and unexposed groups – because they are the same population!



I_{unexp}

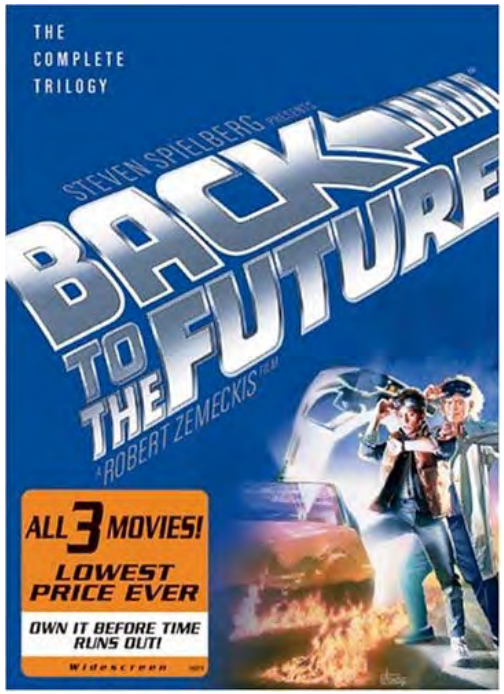
Counterfactual, unexposed cohort

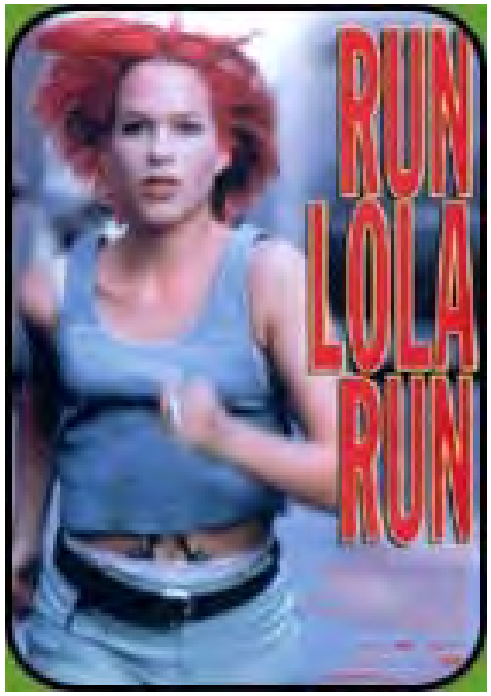
$$RR_{causal} = I_{exp} / I_{unexp}$$

"A causal contrast compares disease frequency under *two* exposure distributions, but in *one* target population during *one* etiologic time period"

Maldonado & Greenland, Int J Epi 2002;31:422-29

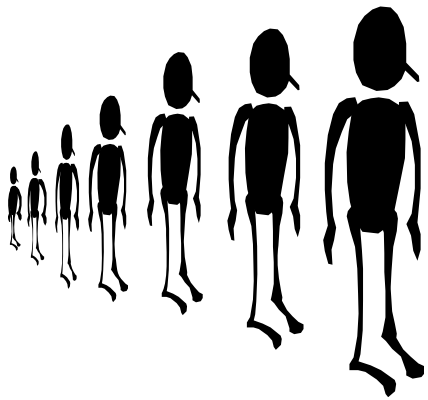
Yes, the counter-factual state
is impossible to observe, unless…





Movies with a ‘counter-factual’ flavor!

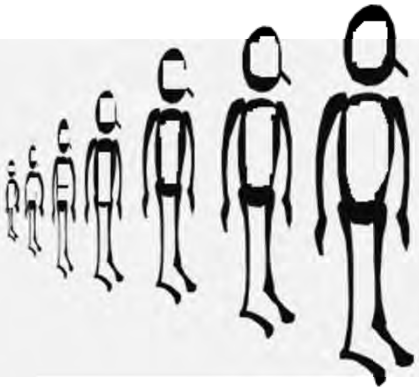
What happens in reality?



Exposed cohort

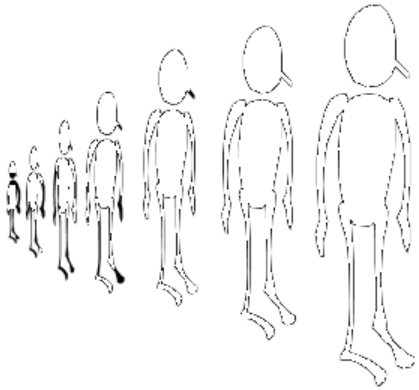
I_{exp}

counterfactual state is not observed (latent)



Counterfactual, unexposed cohort

I_{unexp}



Substitute, unexposed cohort

$I_{substitute}$

A substitute will usually be a population other than the target population during the etiologic time period - INITIAL CONDITIONS MAY BE DIFFERENT

What happens actually?

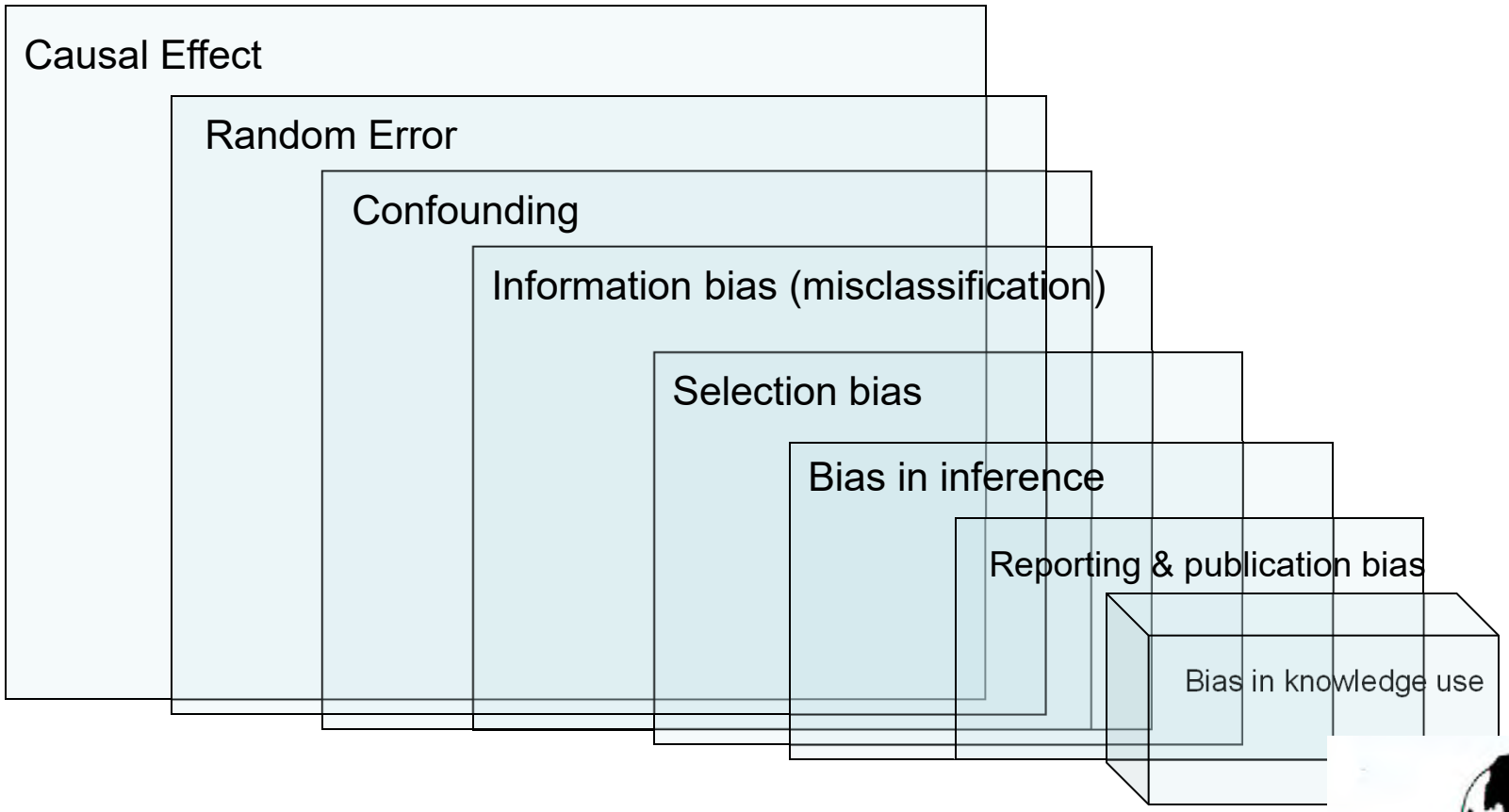
$RR_{causal} = I_{exp} / I_{unexp}$ IDEAL

$RR_{assoc} = I_{exp} / I_{substitute}$ ACTUAL

Chances are…

$RR_{causal} =/= RR_{assoc}$

The best epidemiologic study will be one that captures the causal effect with minimal distortion



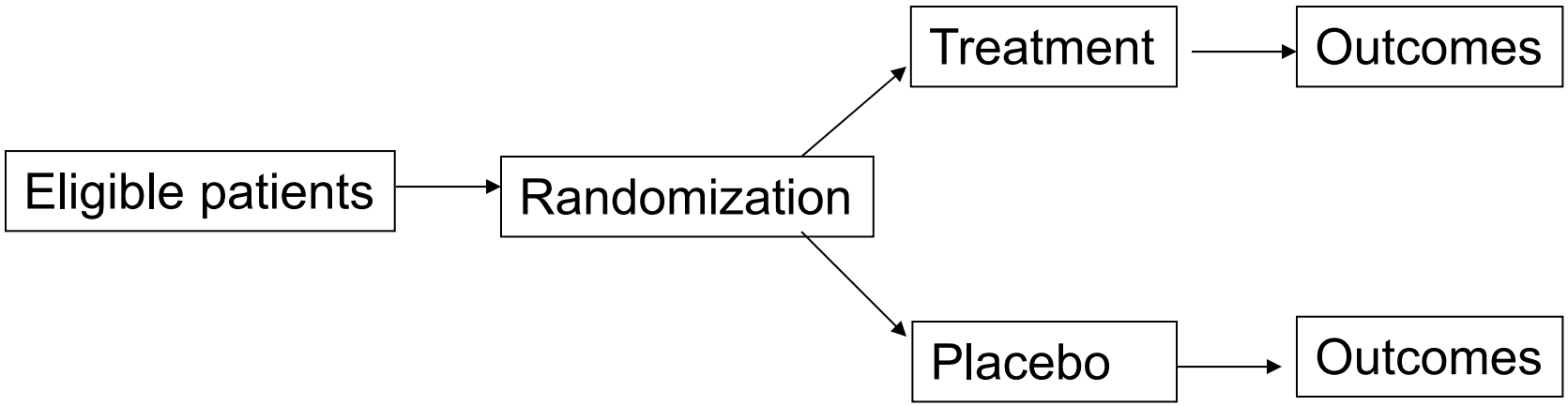
RR_{causal} "truth" → $RR_{association}$

Adapted from: Maclure, M, Schneeweis S. Epidemiology 2001;12:114-122.

Q: What is the Hubble equivalent among epi study designs?

A: the Randomized Controlled Trial (RCT)

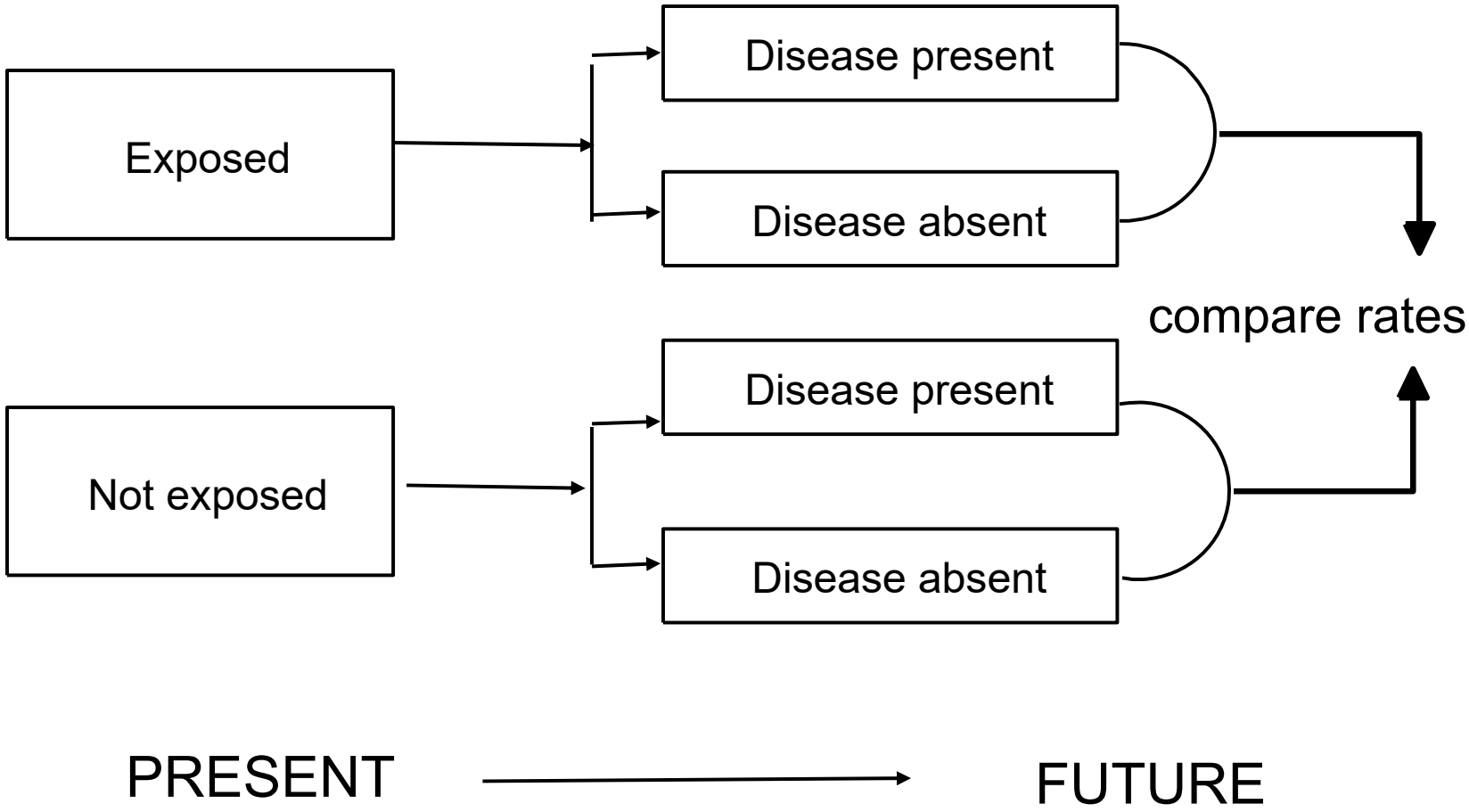
Simulating the counter-factual comparison:
Experimental Studies: RCT



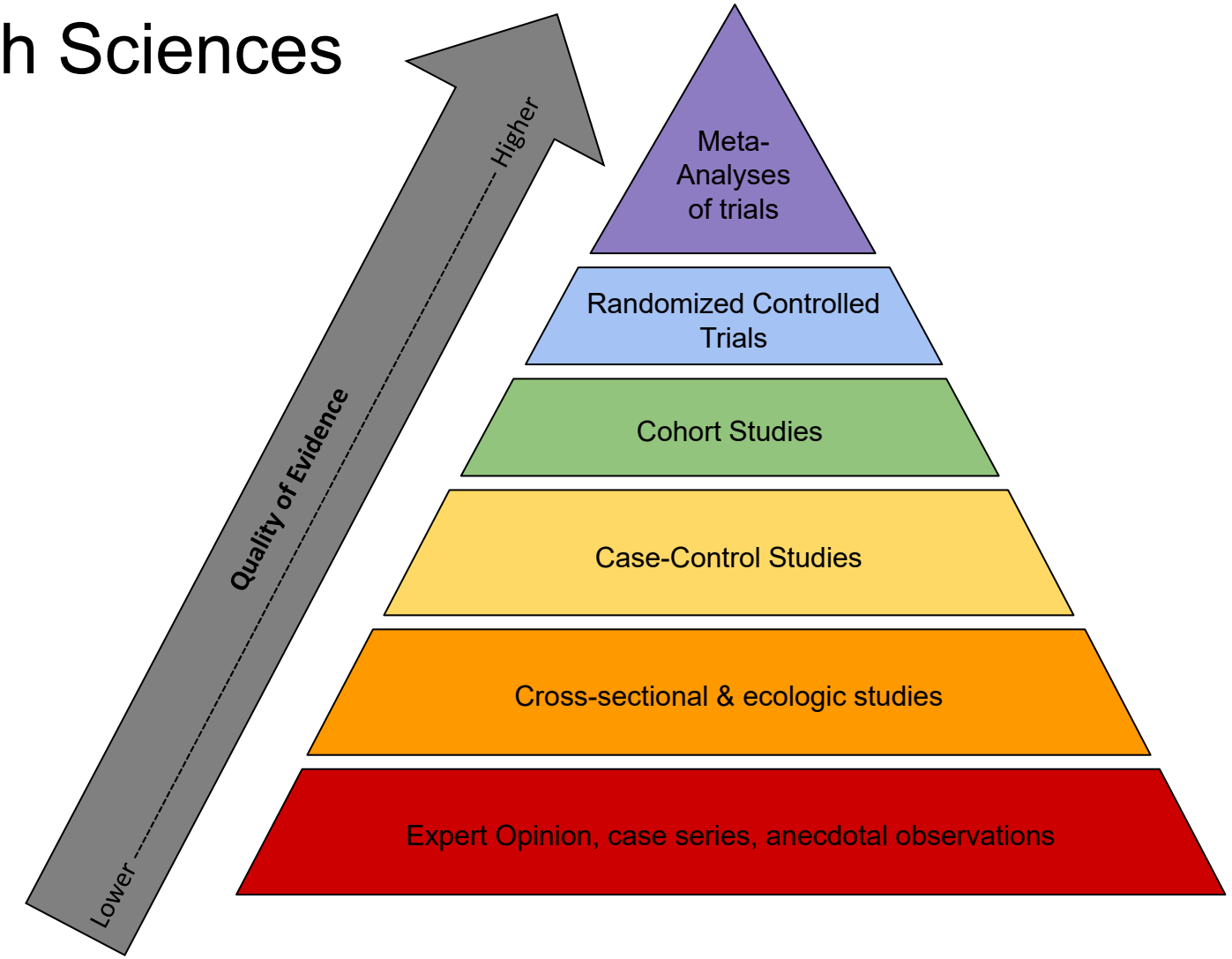
Randomization helps to make the groups “comparable” (i.e. similar initial conditions)

Simulating the counter-factual comparison:
Observational Studies

In observational studies, because exposures are not assigned randomly, attainment of exchangeability is impossible – "initial conditions" are likely to be different and the groups may not be comparable



The Hierarchy of Evidence in the Health Sciences



Adapted from UCI Libraries
https://guides.lib.uci.edu/ebm/pyramid

A general model of causation

- "Causal pie" model by Rothman
  - Sufficient and component cause
  - Strength of effects
  - Interaction among causes
  - Induction period

Rothman, 2002

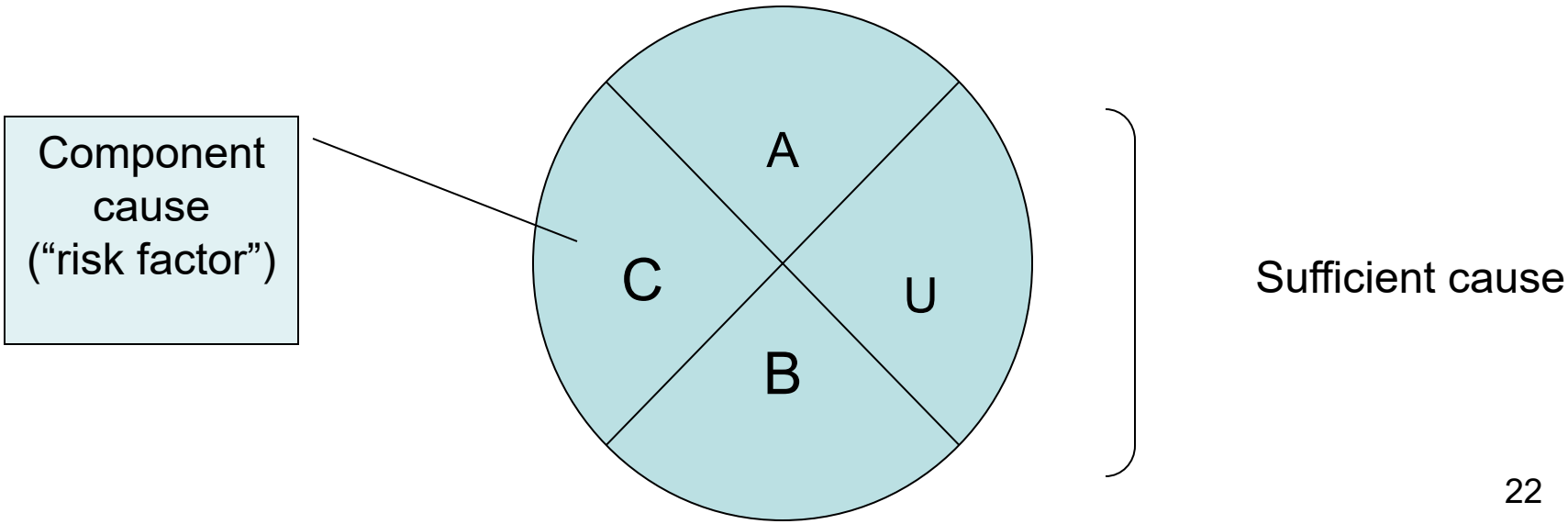
Key premise: disease processes tend to be multifactorial ["multicausality"]

- Very few exposures cause disease entirely by themselves
  - Exposure to measles can cause measles only if somebody is susceptible (e.g. not vaccinated)
  - Development of melanoma among those with high UV light exposure who also have fair skin

Schoenbach, 2000

Sufficient and component causes

- Sufficient causes
  - the whole pie
  - a minimum set of conditions without any one of which the disease would not have occurred
  - often several factors (each factor (slice) is a component cause)
  - Component causes “interact” to produce disease (“biological interaction”) - once all component causes of a sufficient cause are present, that sufficient cause is complete and disease occurs
  - a disease may have several sufficient causes (several pies can produce the same disease)



Rothman, 2002

Sufficient and component causes

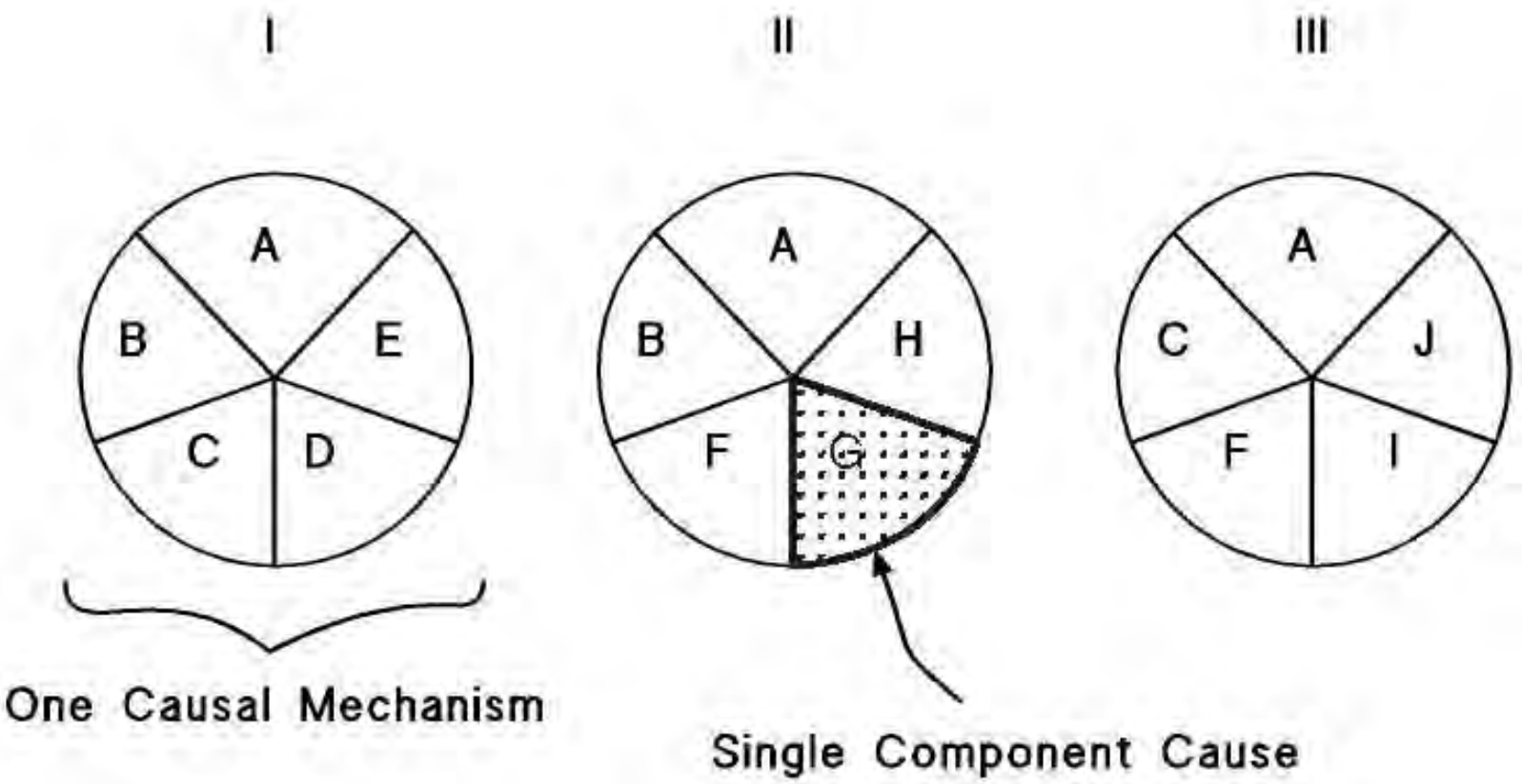
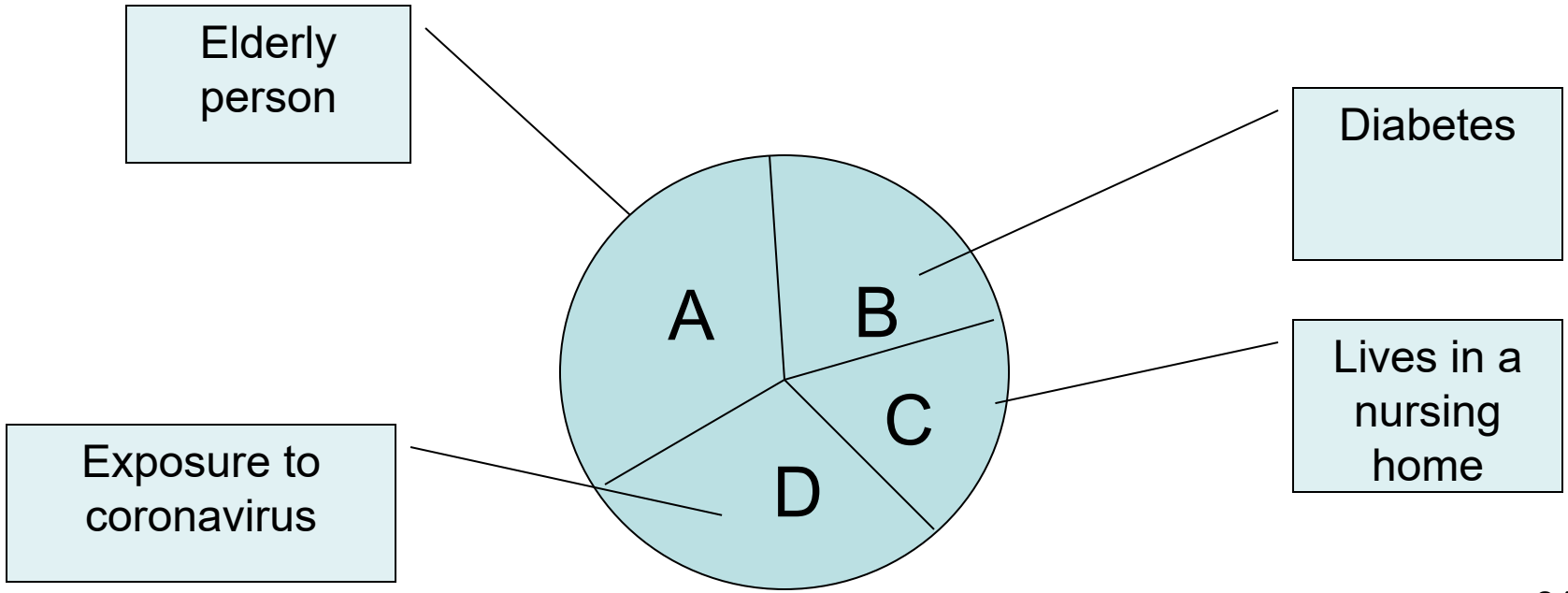


Figure 2–1. Three sufficient causes of a disease.

Rothman, 2002

Example: Covid-19

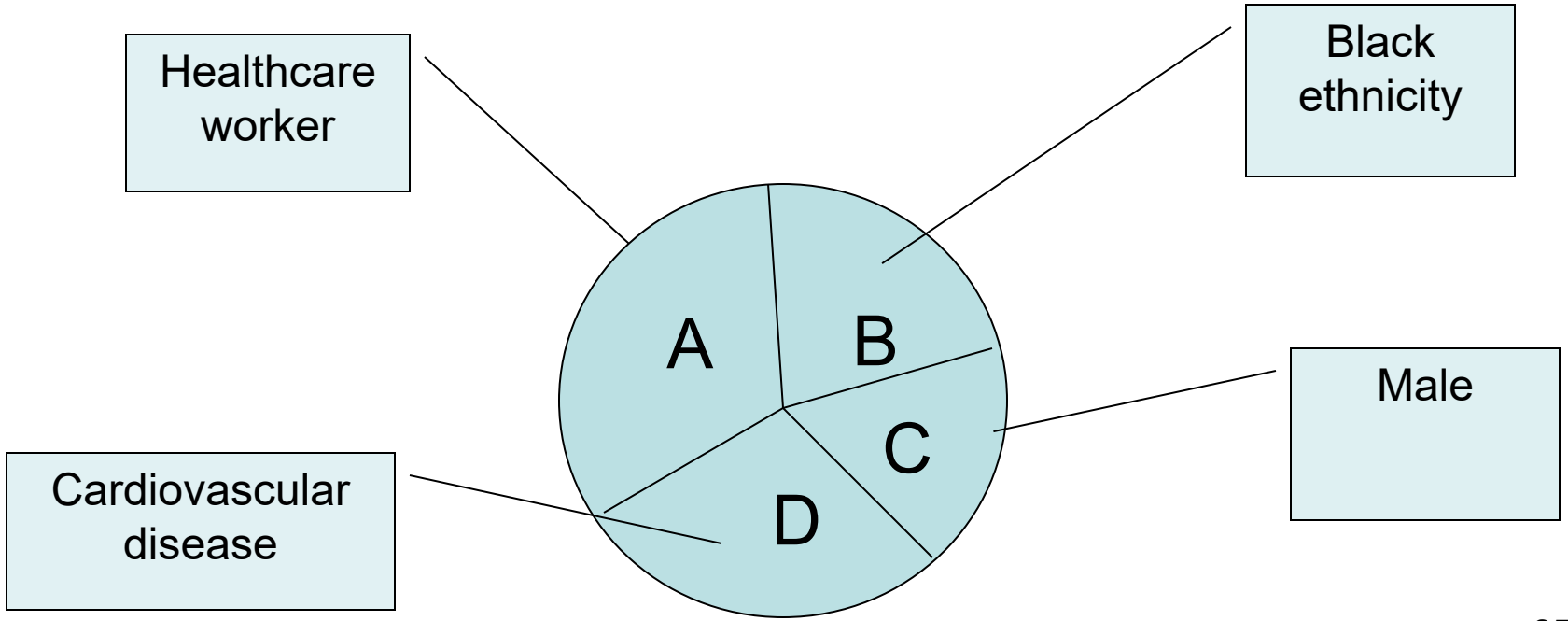
Sufficient cause #1:



Rothman, 2002

Example: Covid-19

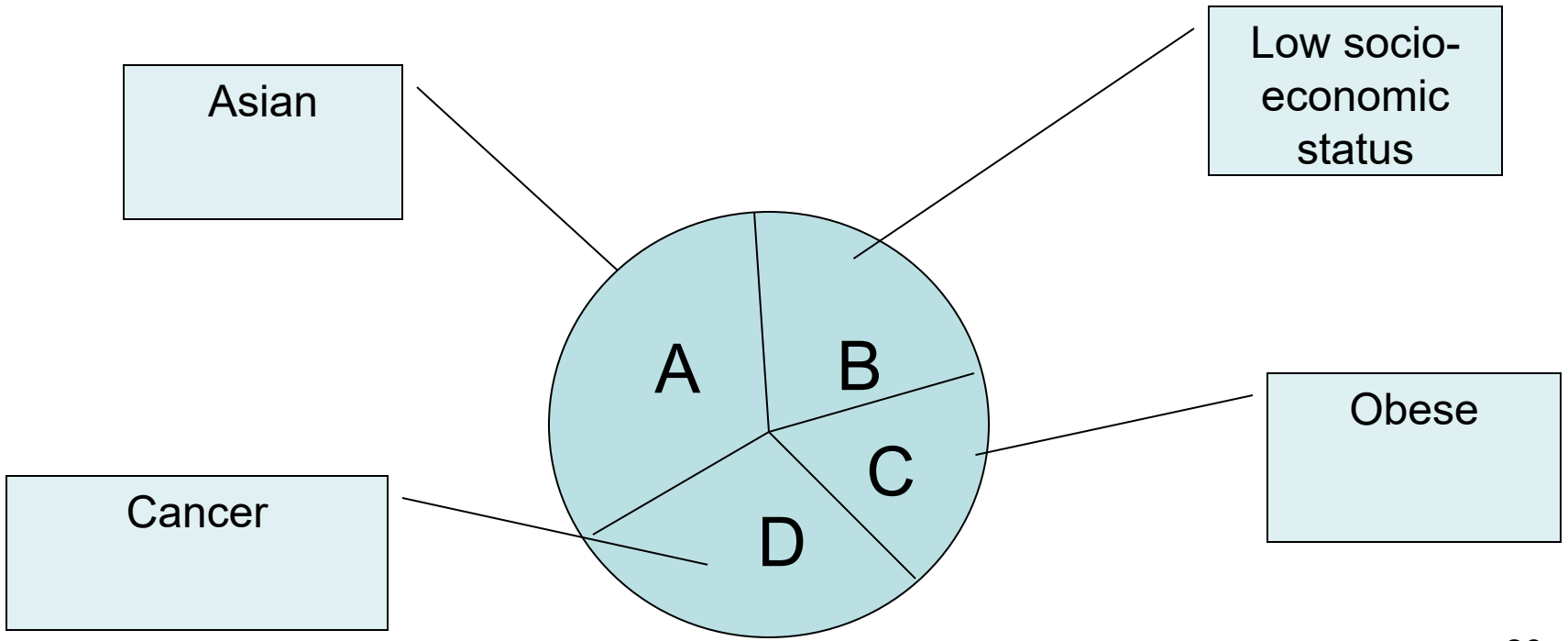
Sufficient cause 2:



Rothman, 2002

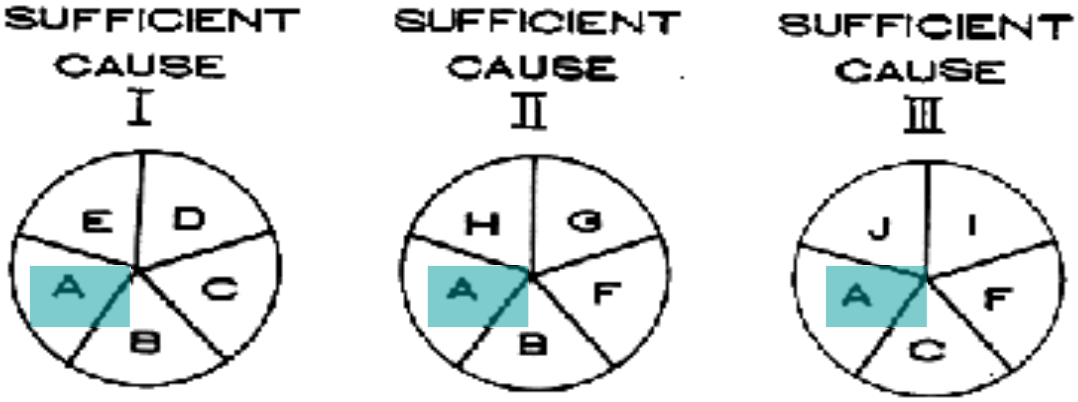
Example: Covid-19

Sufficient cause 3:



Rothman, 2002

Necessary cause



This illustration shows a disease that has 3 sufficient causal complexes, each having 5 component causes.

A is a necessary cause since it appears as a member of each sufficient cause.

B, C, and F are not necessary causes since they fail to appear in all 3 sufficient causes.

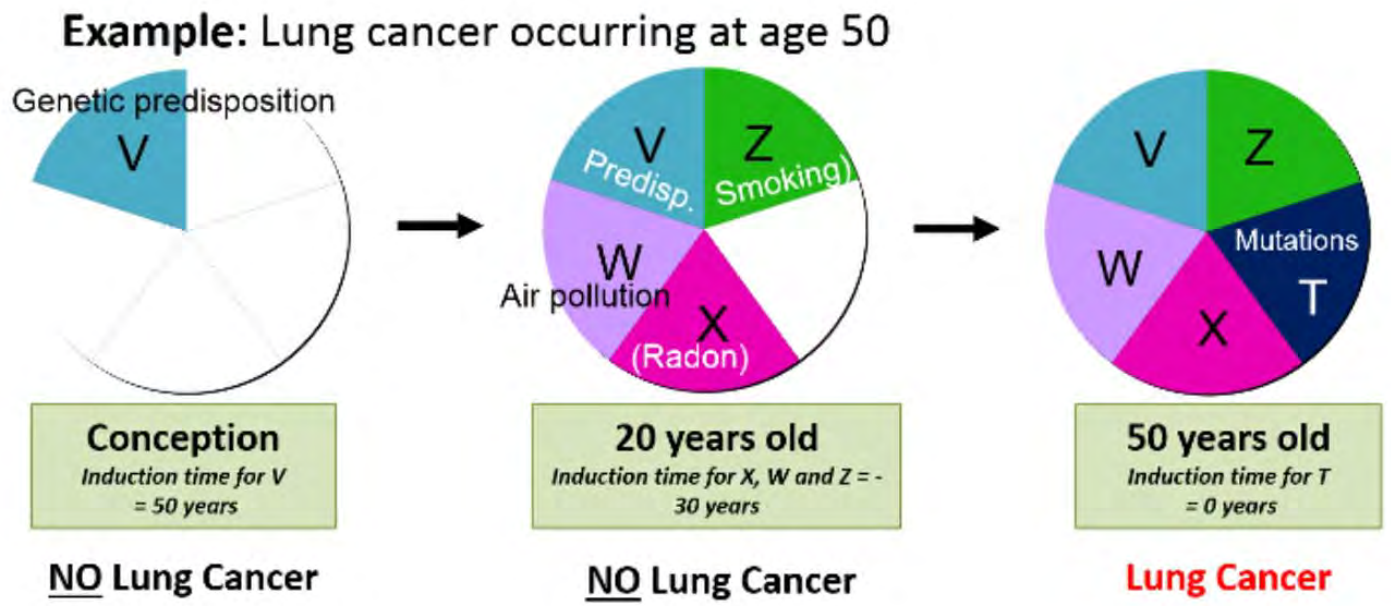
Necessary cause

- most important piece of the pie (without which, disease will not occur)
- must be present for disease to occur
  - Coronavirus is a necessary cause of Covid-19

Rothman, 2002

Timelines

- Completion of a sufficient cause is synonymous with occurrence (although not necessarily diagnosis) of disease:
  - Induction period is the period of time beginning at the action of a component cause and ending when the final component cause acts and the disease occurs (can take years, e.g. cancer)

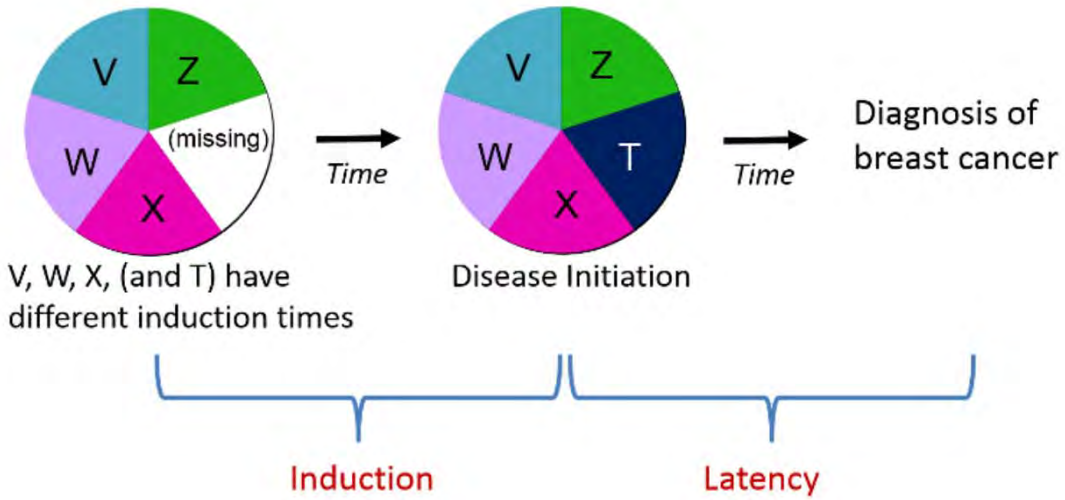


http://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module1A-Populations/PH717-Module1A-Populations8.html

Rothman, 2002

Timelines

- Induction period is the period of time beginning at the action of a component cause and ending when the final component cause acts and the disease occurs (can take years, e.g. cancer)
- Incubation period is used in infectious diseases and is not the same as induction period; incubation period is time from exposure to onset of symptoms
- Latency period: Time from disease to diagnosis



http://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module1A-Populations/PH717-Module1A-Populations8.html

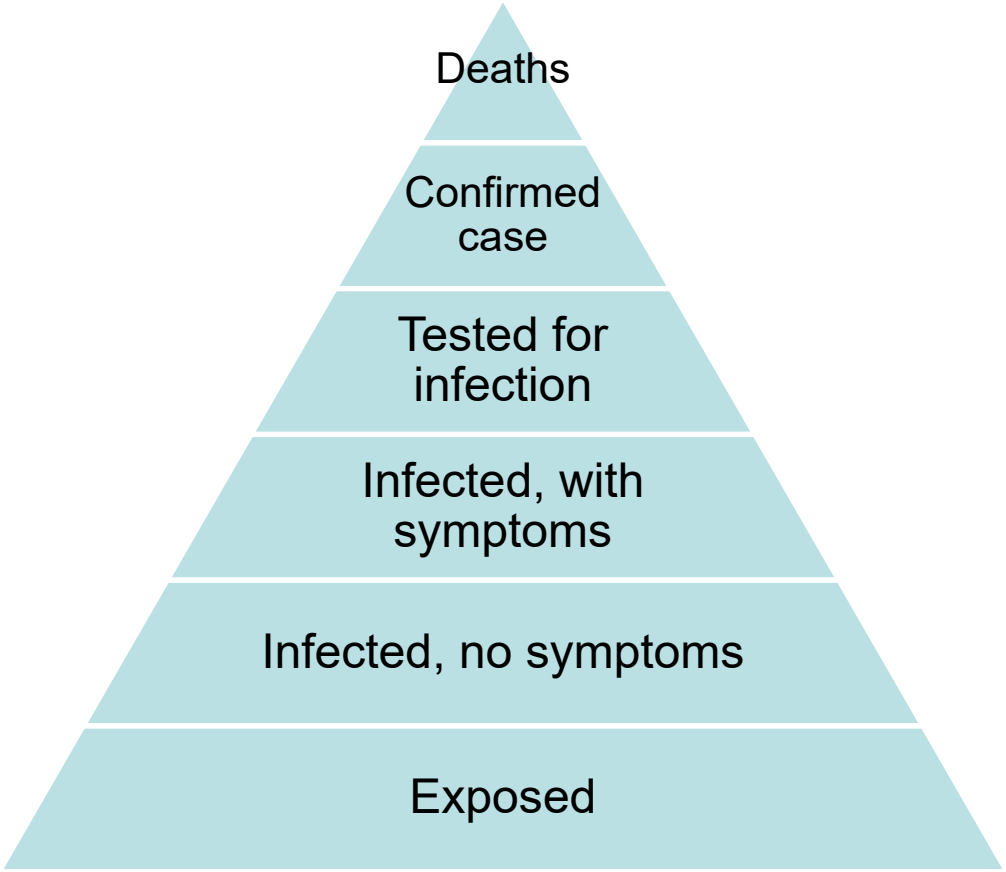
Rothman, 2002

Exposure vs Infection vs Disease

- Exposure = when an individual has encountered a disease causing agent. This is necessary for infection or transmission to take place. However, **exposure need not result in infection**.

- Infection = when an individual has **pathological changes that can be detected**. An infected individual is not necessarily infectious or capable of transmitting the infection.

- Disease = when the body gets altered from normal and functional state to an **abnormal and dysfunctional state**, associated with different signs and symptoms.



With Covid-19, PCR-confirmed cases & deaths are the tip of the iceberg!

https://tinyurl.com/ycqtnqpj

Causal inference using Hill's 'criteria'

Criteria for causation

1. Strength of association
2. Consistency
3. Specificity
4. Temporality
5. Dose-response relationship (gradient)
6. Plausibility
7. Coherence
8. Experimental evidence
9. Analogy

Hill AB. Proc Roy Soc Med 1965

Assessment of the Evidence Suggesting *Helicobacter pylori* as a Causative Agent of Duodenal Ulcers

1. Temporal relationship.
   - *H. pylori* is clearly linked to chronic gastritis. About 11% of chronic gastritis patients will go on to have duodenal ulcers over a 10-year period.
   - In one study of 454 patients who underwent endoscopy 10 years earlier, 34 of 321 patients who had been positive for *H. pylori* (11%) had duodenal ulcer compared with 1 of 133 *H. pylori*-negative patients (0.8%).
2. Strength of the relationship.
   - *H. pylori* is found in at least 90% of patients with duodenal ulcer. In at least one population reported to lack duodenal ulcers, a northern Australian aboriginal tribe that is isolated from other people, it has never been found.
3. Dose-response relationship.
   - Density of *H. pylori* per square millimeter of gastric mucosa is higher in patients with duodenal ulcer than in patients without duodenal ulcer. Also see item 2 above.
4. Replication of the findings.
   - Many of the observations regarding *H. pylori* have been replicated repeatedly.
5. Biologic plausibility.
   - Although originally it was difficult to envision a bacterium that infects the stomach antrum causing ulcers in the duodenum, it is now recognized that *H. pylori* has binding sites on antral cells and can follow these cells into the duodenum.
   - *H. pylori* also induces mediators of inflammation.
   - *H. pylori*-infected mucosa is weakened and is susceptible to the damaging effects of acid.
6. Consideration of alternate explanations.
   - Data suggest that smoking can increase the risk of duodenal ulcer in *H. pylori*-infected patients but is not a risk factor in patients in whom *H. pylori* has been eradicated.
7. Cessation of exposure.
   - Eradication of *H. pylori* heals duodenal ulcers at the same rate as histamine receptor antagonists.
   - Long-term ulcer recurrence rates were zero after *H. pylori* was eradicated using triple-antimicrobial therapy, compared with a 60% to 80% relapse rate often found in patients with duodenal ulcers treated with histamine receptor antagonists.
8. Specificity of the association.
   - Prevalence of *H. pylori* in patients with duodenal ulcers in 90% to 100%. However, it is found in some patients with gastric ulcer and even in asymptomatic individuals.
9. Consistency with other knowledge.
   - Prevalence of *H. pylori* infection is the same in men as in women. The incidence of duodenal ulcer, which in earlier years was believed to be higher in men than in women, has been equal in recent years.
   - The prevalence of ulcer disease in believed to have peaked in the latter part of the 19th century, and the prevalence of *H. pylori* may have been much higher at that time because of poor living conditions. This reasoning is also based on observations today that the prevalence of *H. pylori* is much higher in developing countries.

Data from Megraud F, Lamouliatte H: *Helicobactor pylori* and duodenal ulcer: Evidence suggesting causation. Dig Dis Sci 37:769-772, 1992; and DeCross AJ, Marshall BJ: The role of *Helicobactor pylori* in acid-peptic disease. Am J Med Sci 306:381-391, 1993.

Single studies are never enough to make a causal inference!

Correlation ≠ Causation

- **Correlations** between variables indicate that one variable changes with the other
- **Causation** proposes that one variable causes a change in another

COVID-19 and BCG example:

Although there is a correlation, there are many concerns which prevent causal inference:

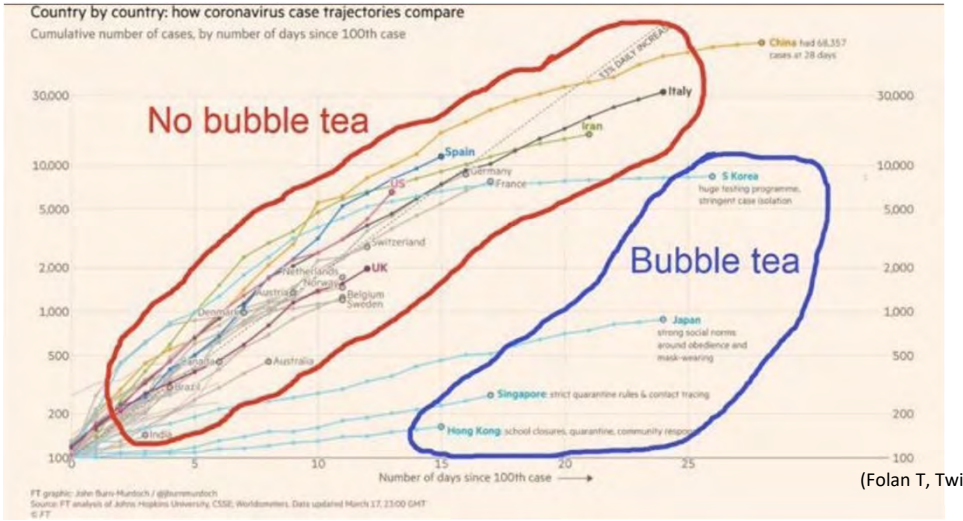
-Variable testing rates

-Confounding by age & other factors

-Selection bias (who gets BCG?), etc

-Ecological correlation may not hold at individual level

Even better example: COVID-19 and Bubble Tea



(Folan T, Twitter)

https://tinyurl.com/t2jmyf6