epidemiology: the big picture

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

McGill

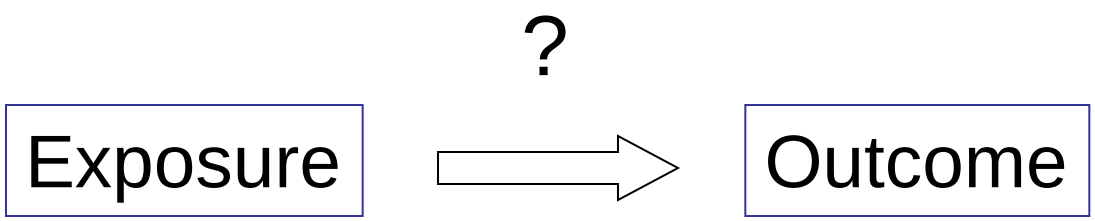
Why epidemiology?

- We are engaged in healthcare and health research
- To effectively practice medicine and public health, we need evidence/knowledge on 3 fundamental types of professional knowing “gnosis”:

Dia-gnosis	Etio-gnosis	Pro-gnosis	For individual (Clinical Medicine)
Dia-gnosis	Etio-gnosis	Pro-gnosis	For community (Public and community health)

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)



Causality: outline

- What is a cause?
- What is a causal effect? (“counterfactuals”)
- A general model of causation (“causal pie model”)
- Philosophy of scientific inference
  - Inductivism
  - Refutationism
  - Conjecture and refutation
- Causal inference in epidemiology
  - Causal criteria (Hill’s)
  - Testing competing theories (“Strong Inference”)

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.”

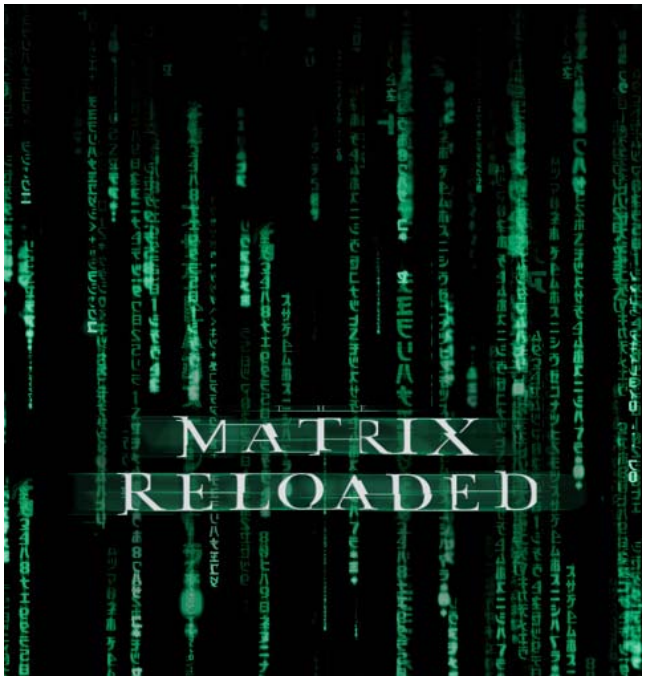
  » [Rothman & Greenland, 1998]

- “The primary ‘knowledge object’ of epidemiology as a scientific discipline are causes of health-related events in populations.” [Porta, IEA Dictionary, 2008]

Importance of Causality: Matrix

- “You see, there is only one constant, one universal, it is the only real truth: causality. Action. Reaction. Cause and effect.
- Causality. There is no escape from it, we are forever slaves to it. Our only hope, our only peace is to understand it, to understand the `why.' ` Why' is what separates us from them, you from me. `Why' is the only real social power, without it you are powerless.”

  -The Merovingian

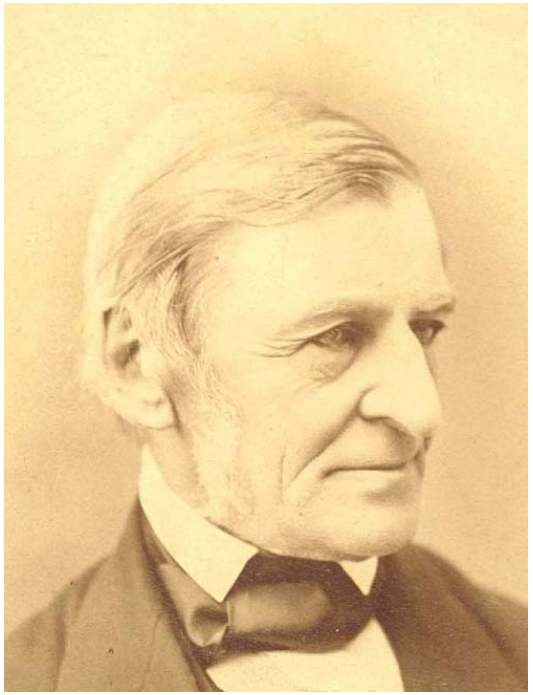


Emerson's view on causality

"Shallow men believe in luck. Strong men believe in cause and effect."

Ralph Waldo Emerson

1803 –1882



Cause and effect?

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
    - negative = the absence of exposure (e.g. vaccination)

- Should 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

Direction of causality: does overeating cause obesity?

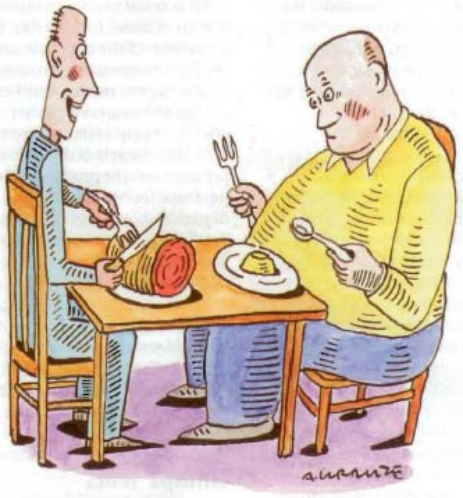
The great diet delusion

It's time for nutrition researchers and health authorities to wake up to the fact that people do not get fat simply because they overeat, says Gary Taubes

FOR the past century, the advice to the overweight and obese has remained remarkably consistent: consume fewer calories than you expend and you will lose weight. This prescription seems eminently reasonable. The only problem is that it doesn't seem to work. Neither eating less nor moving more reverses the course of obesity in any but the rarest cases.

This contradiction has given us a catalogue of clinical literature almost mind-boggling in its internal inconsistency. "Dietary therapy remains the cornerstone of [obesity] treatment and the reduction of energy intake continues to be the basis of successful weight reduction programs," observes *The Handbook of Obesity*, a textbook edited by George Bray, Claude Bouchard and W. P. T. James, three of the most respected names in obesity research, and first published in 1998. It then goes on to acknowledge that the results of such therapy "are known to be poor and not long-lasting".

In truth, the very idea that such advice might benefit obese people borders on the nonsensical, presupposing as it does that they are either unconcerned about their weight, ignorant, or stubbornly unwilling to do anything about it. None of these notions has a shred of evidence to support it, yet health authorities still repeat their mantra: obesity is caused by overeating; eating less is the cure. Any attempt to argue otherwise is treated as quackery.



The institutionalised conviction that we get fat simply because we overeat is based on the kind of fallacious reasoning that would lead to a failing grade in a high-school logic class. The first law of thermodynamics tells us that energy is neither created nor destroyed, so the calories we consume must be either stored, expended or excreted. If we are getting fatter, we must be taking in more energy than we are giving out: we are overeating.

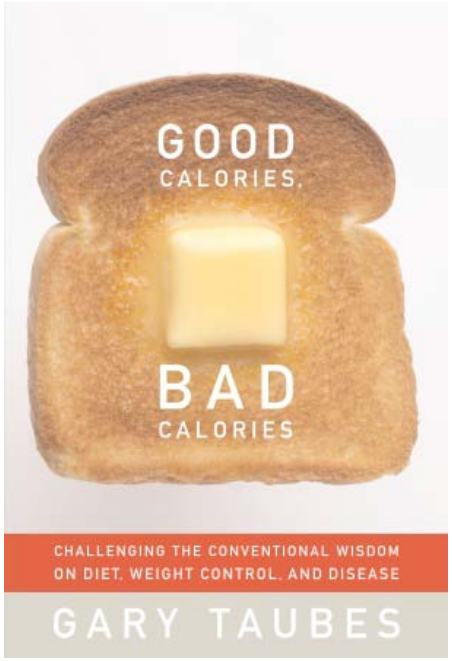
> "They have created a field of clinical medicine that functions more like a religion than a science"

to the second world war when they discussed the causes of obesity, which they considered very obviously a hormonal and genetic disorder. When children go through growth spurts, they are likely to eat voraciously – to overeat. Indeed, they have to do this to fuel their growing bodies. They do not grow because they overeat; they overeat because they are growing. The growth is induced by hormones, specifically growth hormone.

Just as you can starve a fat person and induce them to lose weight, you can starve a child and prevent them from growing. Neither implies that overeating was the root cause of their getting fat or growing bigger.

There is considerable evidence that the obesity epidemic is caused by a hormonal phenomenon, specifically by the consumption of refined carbohydrates, starches and sugars, all of which prompt (sooner or later) excessive insulin secretion. Insulin is the primary regulator of fat storage. When insulin levels are elevated, fat accumulates in our body tissue; when they fall, fat is released and we use it for fuel. By stimulating insulin secretion, carbohydrates make us fat; by driving us to accumulate fat, they increase hunger and decrease the energy we expend in metabolism and physical activity. In short, obesity is caused not by overeating or sedentary behaviour, but by hormonal malfunctioning triggered by the consumption of particular types of carbohydrate-containing foods.

Obesity researchers, nutritionists and health authorities have refused to contemplate this scenario, partly because it would imply that diet-book doctors advocating carbohydrate-restricted diets – Robert Atkins et



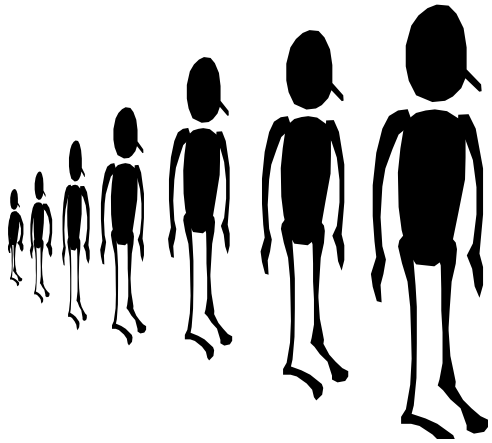
Taubes G
New Scientist
2008

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

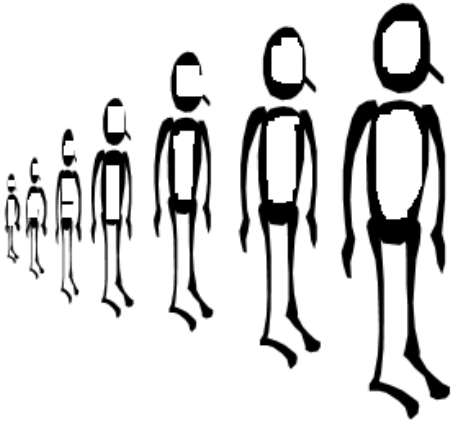
Ideal counterfactual comparison to determine causal effects



I_{exp}

Exposed cohort

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



I_{unexp}

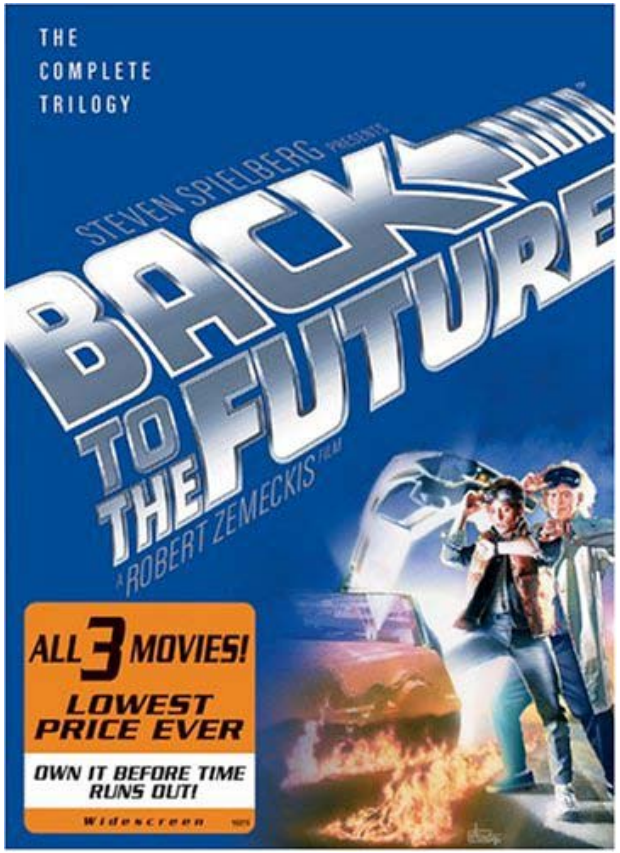
Counterfactual, unexposed cohort

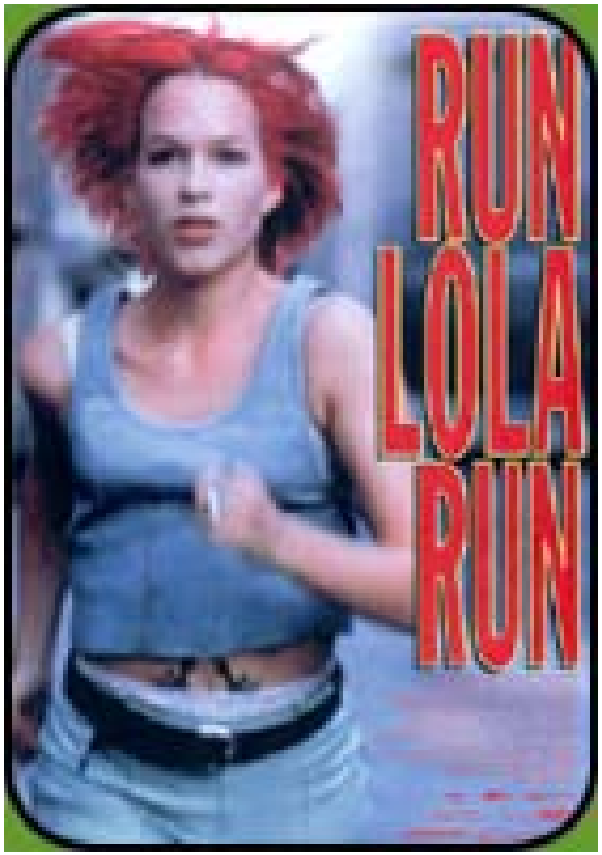
$$RR_{causal} = I_{exp} / I_{unexp}$$

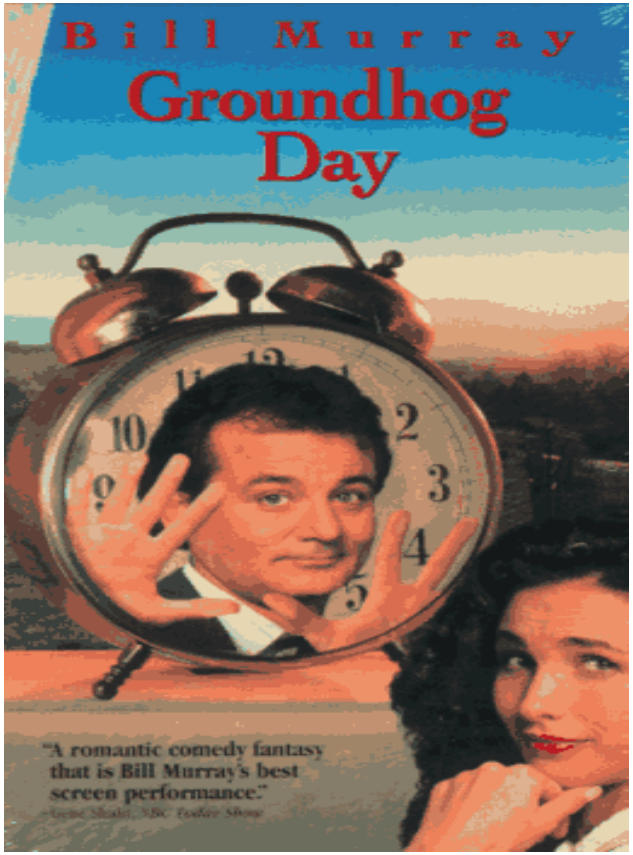
"A causal contrast compares disease frequency under *two* exposure distributions, but in *one* target population during *one* etiologic time period"

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

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

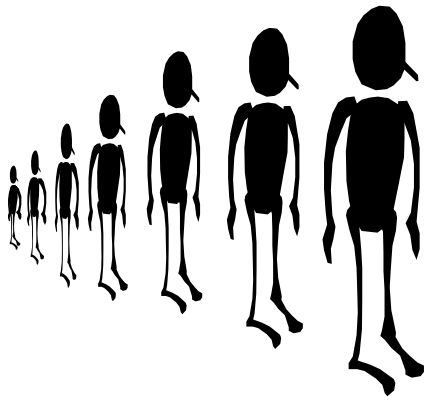






Movies with a ‘counter-factual’ flavor!

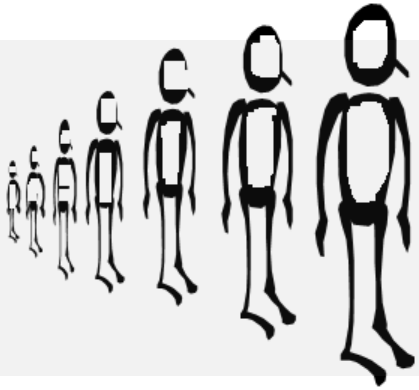
What happens in reality?



Exposed cohort

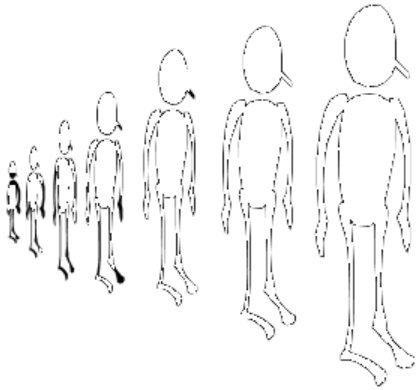
I_{exp}

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?

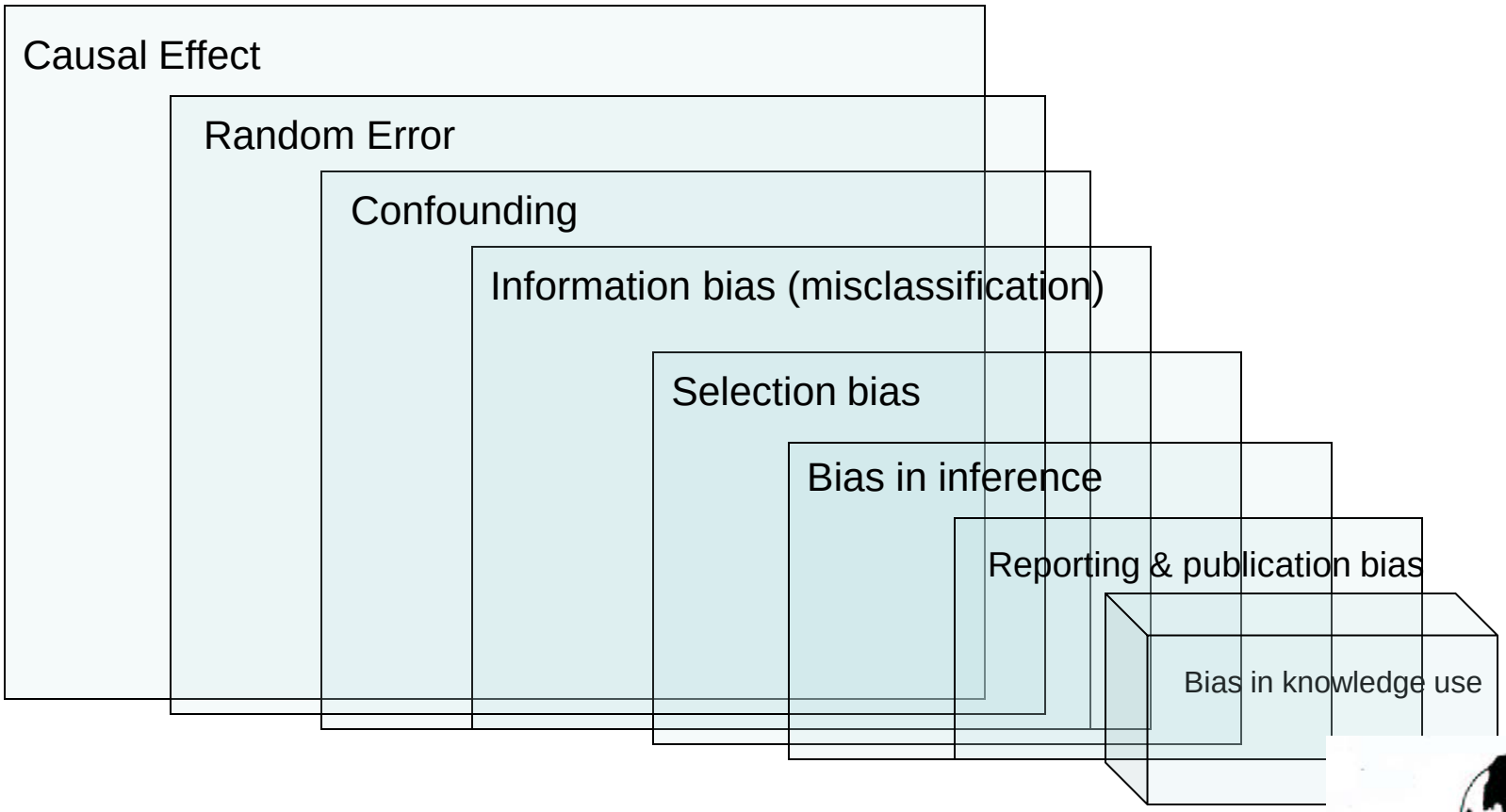
$RR_{causal} = I_{exp} / I_{unexp}$ IDEAL

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

Chances are…

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

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



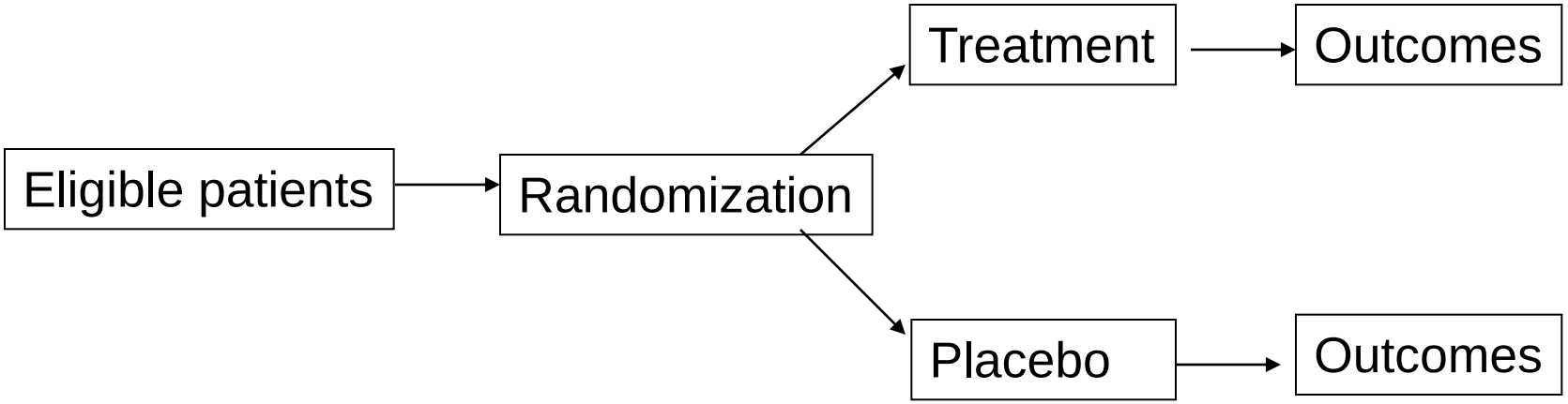
RR_{causal} "truth" → $RR_{association}$

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

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

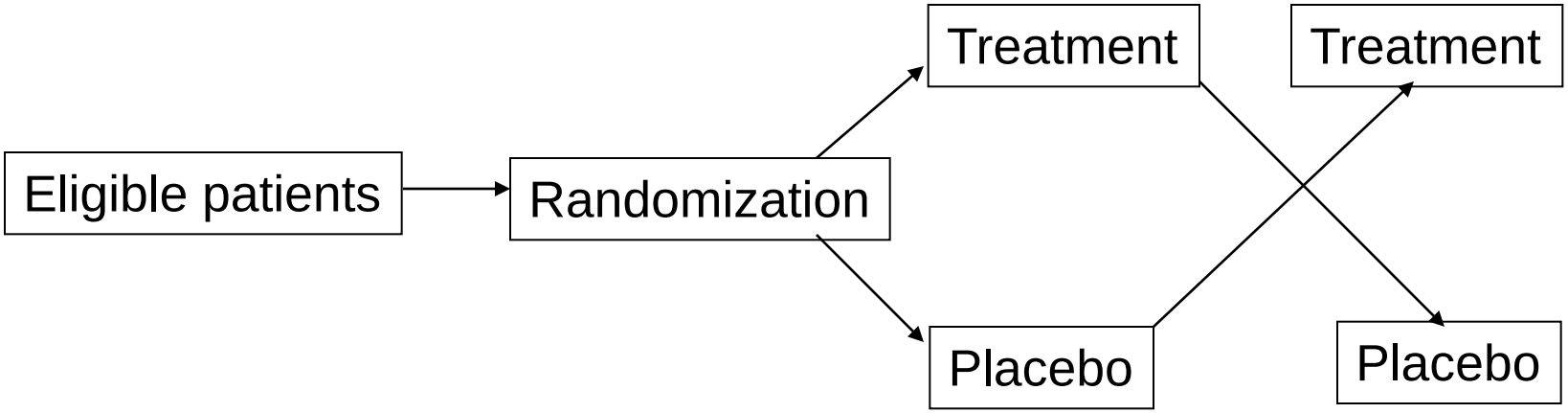
A: the Randomized Controlled Trial (RCT)

Simulating the counter-factual comparison:
Experimental Studies: RCT



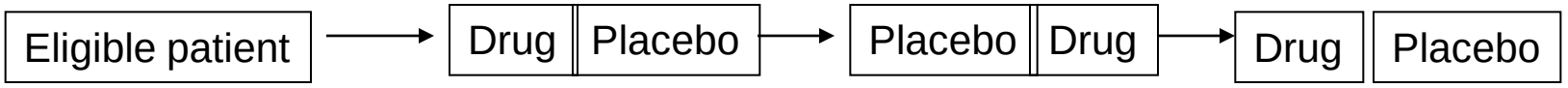
Randomization helps to make the groups “comparable” (i.e. similar initial conditions)

Simulating the counter-factual comparison:
Experimental Studies: Cross-over trials



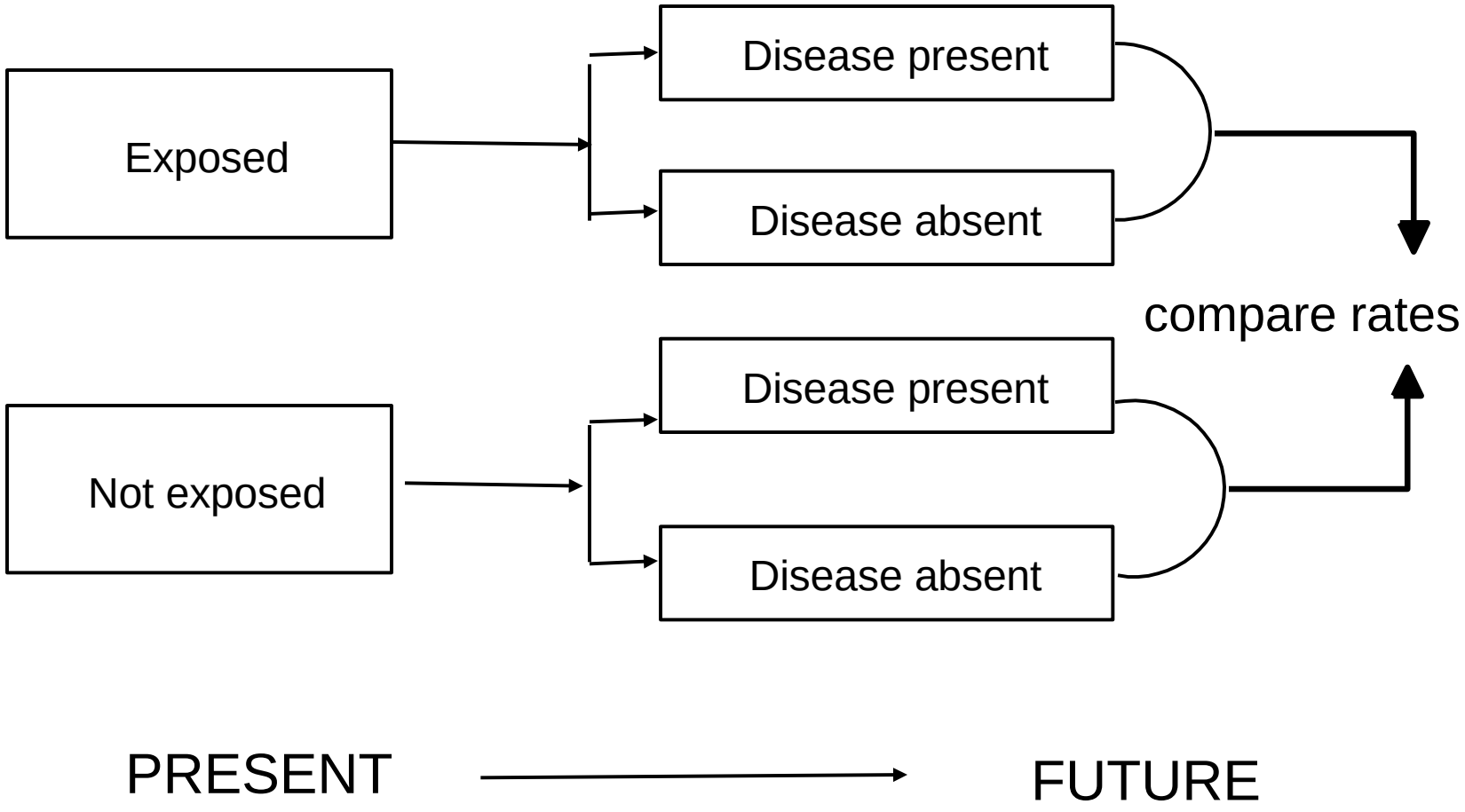
Although cross-over trials come close to the ideal of counterfactual comparison, they do not achieve it because a person can be in only one study group at a time; variability in other exposures across time periods can still introduce confounding (Rothman, 2002)

N-of-1 Trial



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



Hierarchy of evidence

Grade of Recommendation	Level of Evidence	Therapy/Prevention, Aetiology/Harm	Prognosis	Diagnosis	Economic analysis
A	1a	SR (with homogeneity*) of RCTs	SR (with homogeneity*) of inception cohort studies; or a CPG† validated on a test set.	SR (with homogeneity*) of Level 1 diagnostic studies; or a CPG validated on a test set.	SR (with homogeneity*) of Level 1 economic studies
	1b	Individual RCT (with narrow Confidence Interval‡)	Individual inception cohort study with ≥ 80% follow-up	Independent blind comparison of an appropriate spectrum of consecutive patients, all of whom have undergone both the diagnostic test and the reference standard.	Analysis comparing all (critically-validated) alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.
	1c	All or none§	All or none case-series¤	Absolute SpPins and SnNouts††	Clearly as good or better‡‡, but cheaper. Clearly as bad or worse but more expensive. Clearly better or worse at the same cost.
B	2a	SR (with homogeneity*) of cohort studies	SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs.	SR (with homogeneity*) of Level ≥2 diagnostic studies	SR (with homogeneity*) of Level ≥2 economic studies
	2b	Individual cohort study (including low quality RCT; e.g., <80% follow-up)	Retrospective cohort study or follow-up of untreated control patients in an RCT; or CPG not validated in a test set.	Independent blind comparison but either in non-consecutive patients, or confined to a narrow spectrum of study individuals (or both), all of whom have undergone both the diagnostic test and the reference standard; or a diagnostic CPG not validated in a test set.	Analysis comparing a limited number of alternative outcomes against appropriate cost measurement, and including a sensitivity analysis incorporating clinically sensible variations in important variables.
	2c	"Outcomes" Research	"Outcomes" Research		
	3a	SR (with homogeneity*) of case-control studies			
	3b	Individual Case-Control Study		Independent blind comparison of an appropriate spectrum, but the reference standard was not applied to all study patients	Analysis without accurate cost measurement, but including a sensitivity analysis incorporating clinically sensible variations in important variables.
C	4	Case-series (and poor quality cohort and case-control studies§§)	Case-series (and poor quality prognostic cohort studies ¤¤)	Reference standard was not applied independently or not applied blindly	Analysis with no sensitivity analysis
D	5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles"	Expert opinion without explicit critical appraisal, or based on economic theory

Hierarchy of evidence

TABLE 1A–1

A Hierarchy of Strength of Evidence for Treatment Decisions

- N of 1 randomized controlled trial
- Systematic reviews of randomized trials
- Single randomized trial
- Systematic review of observational studies addressing patient-important outcomes
- Single observational study addressing patient-important outcomes
- Physiologic studies (studies of blood pressure, cardiac output, exercise capacity, bone density, and so forth)
- Unsystematic clinical observations

Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice AMA Press, 2002 [Editors Guyatt & Rennie]

A general model of causation

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

Rothman, 2002

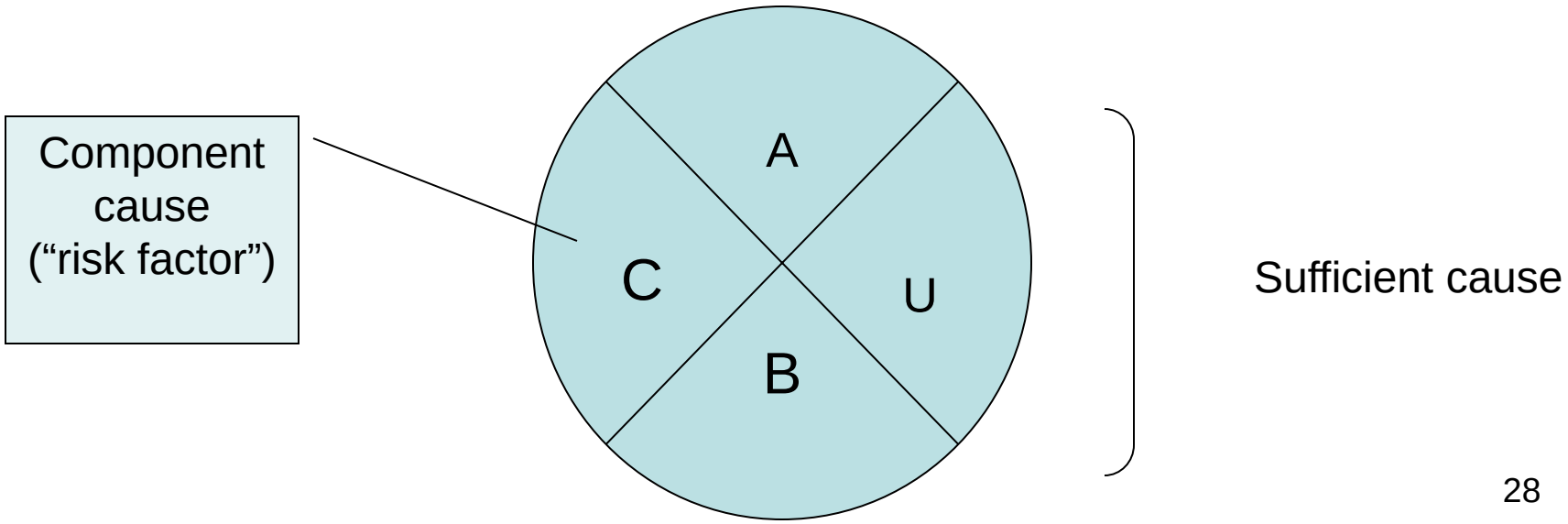
Key premise: disease processes tend to be multifactorial ["multicausality"]

- 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

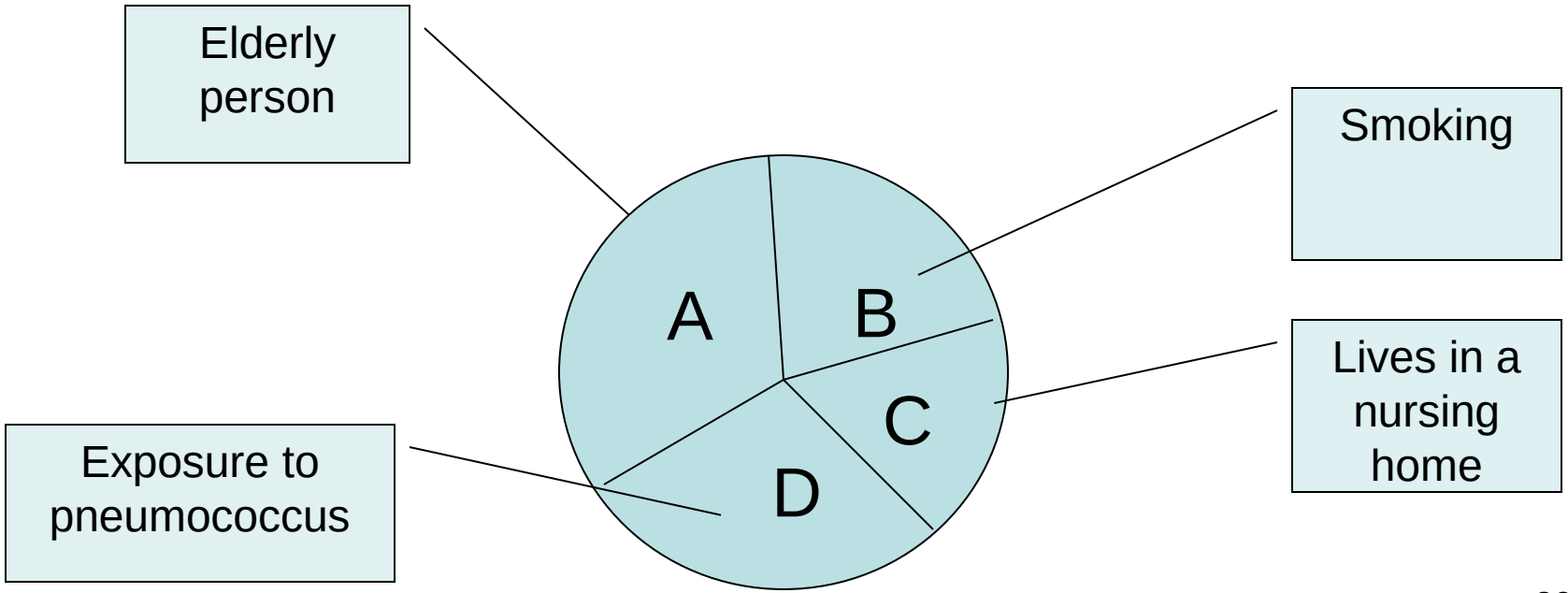
- Sufficient causes
  - the whole pie
  - a minimum set of conditions without any one of which the disease would not have occurred
  - not usually a single factor, often several (each factor (slice) is a component cause)
  - Component causes “interact” to produce disease
  - a disease may have several sufficient causes (several pies can produce the same disease)



Rothman, 2002

Example: Pneumonia

One of many possible sufficient causes:



Rothman, 2002

Sufficient and component causes

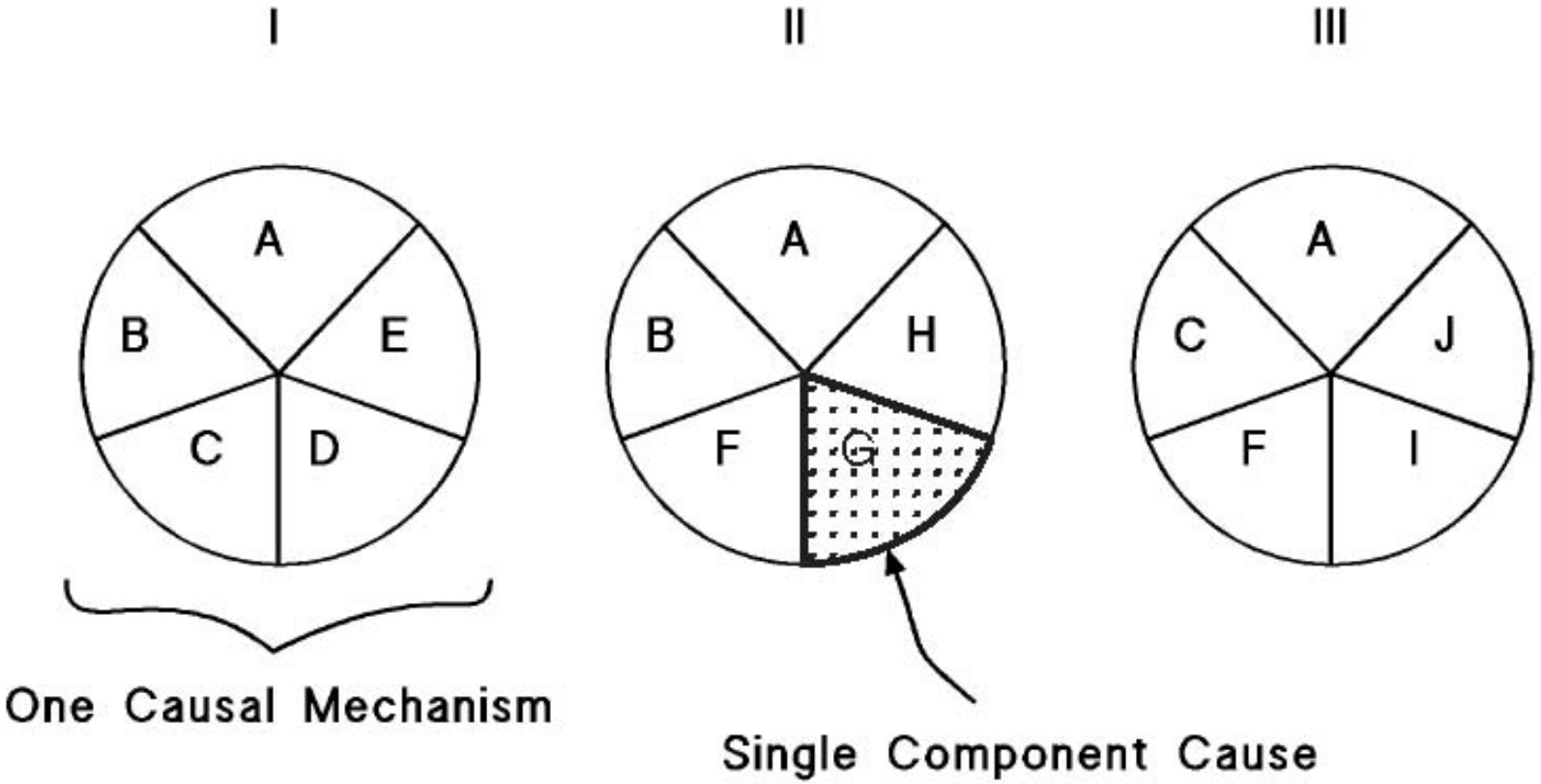
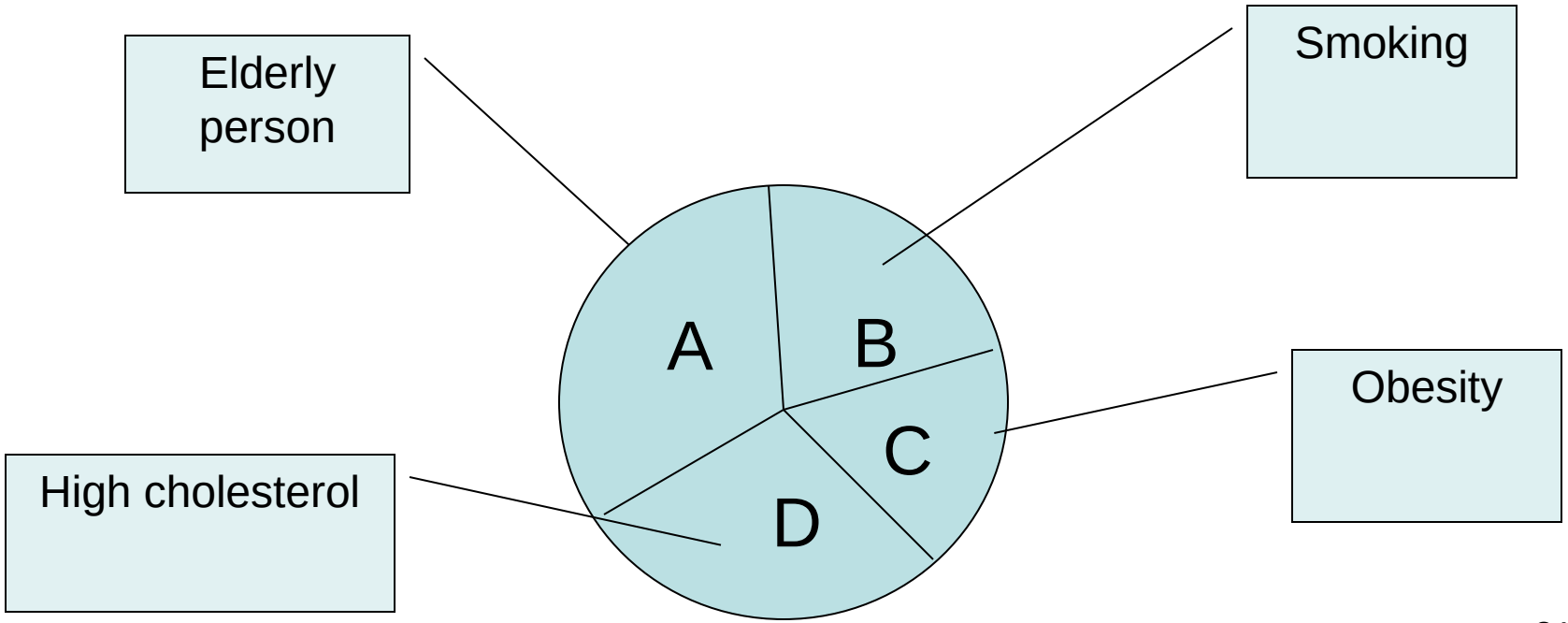


Figure 2–1. Three sufficient causes of a disease.

Rothman, 2002

Example: Coronary artery disease

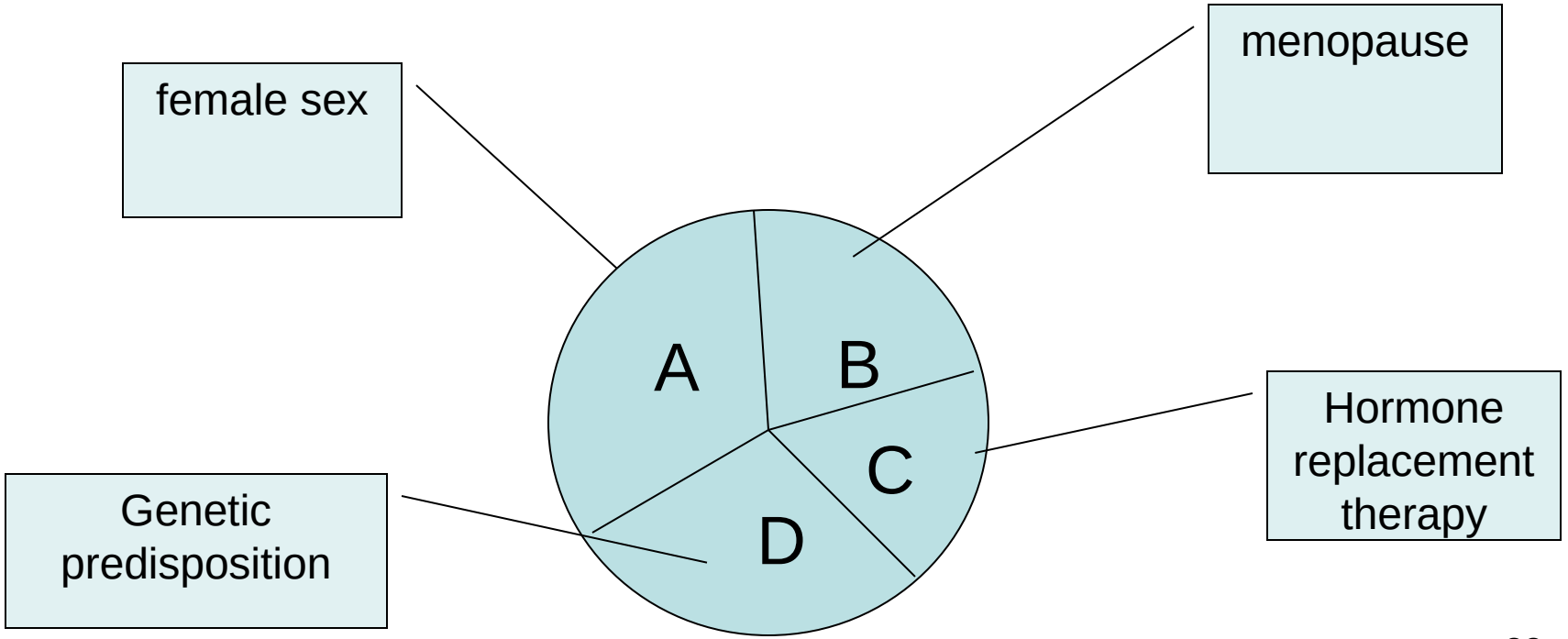
Model 1:



Rothman, 2002

Example: Coronary artery disease

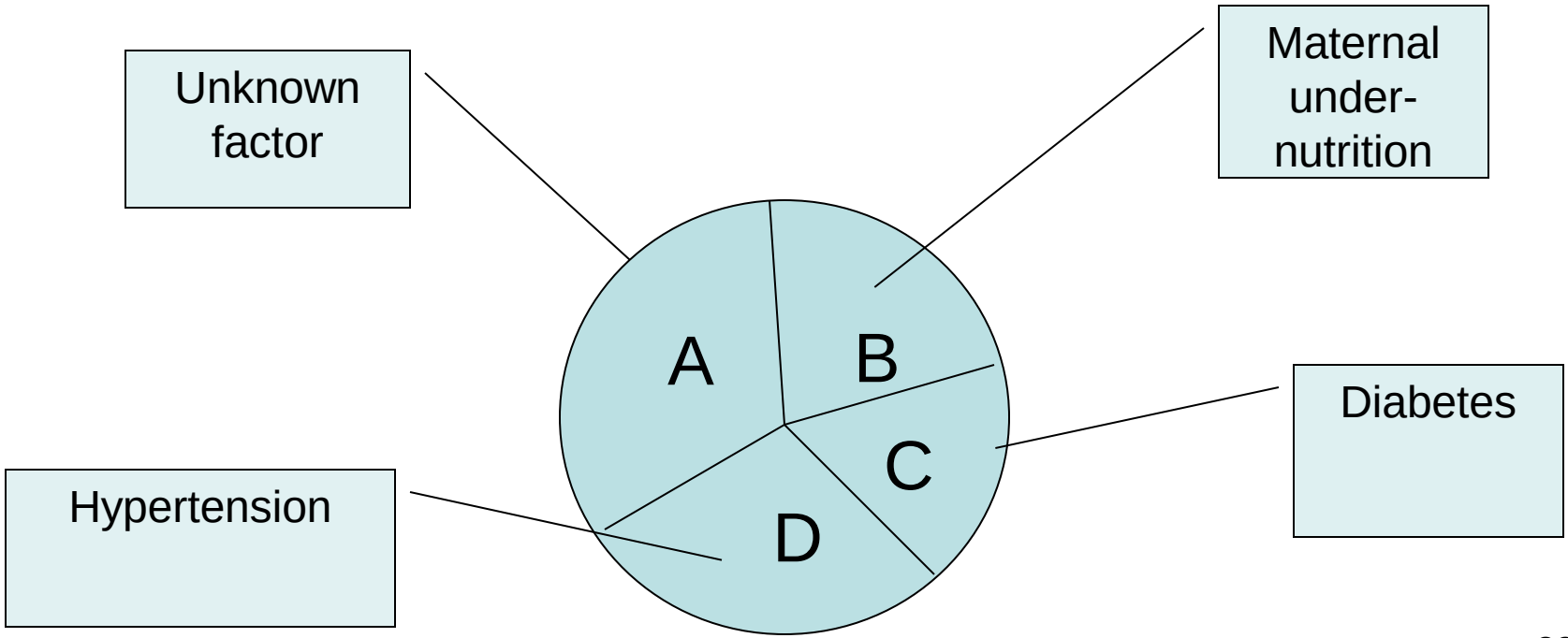
Model 2:



Rothman, 2002

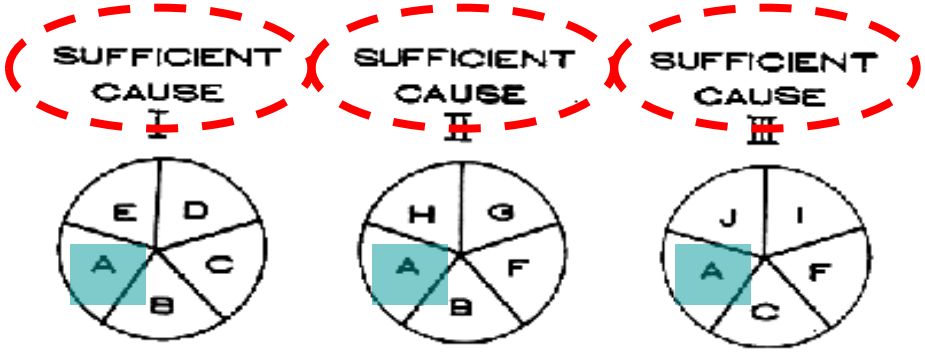
Example: Coronary artery disease

Model 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

- the most important piece of the pie (without which, disease will not occur for that individual)
- must be present for disease to occur
  - HIV is a necessary cause of AIDS

Rothman, 2002

Does HIV really cause AIDS?

OPEN ACCESS Freely available online

PLoS MEDICINE

Policy Forum

HIV Denial in the Internet Era

Tara C. Smith*, Steven P. Novella

It may seem remarkable that, 23 years after the identification of the human immunodeficiency virus (HIV), there is still denial that the virus is the cause of acquired immune deficiency syndrome (AIDS). This denial was highlighted on an international level in 2000, when South African president Thabo Mbeki convened a group of panelists to discuss the cause of AIDS, acknowledging that he remained unconvinced that HIV was the cause [1]. His ideas were derived at least partly from material he found on the Internet [2]. Though Mbeki agreed later that year to step back from the debate [3], he subsequently suggested a re-analysis of health spending with a decreased emphasis on HIV/AIDS [4].

HIV denial has taken root in the general population and has shown its potential to frustrate public education efforts and adversely affect public funding for AIDS research and

doi:10.1371/journal.pmed.0040256.g001

Example of a typical slogan from an HIV denialist group

that HIV does not cause AIDS at concerts [6], and it lists the HIV denial group "Alive and Well" as a worthy cause on its Web site (http://www.

to many other forms of popular denial, including denial of evolution, mental illness, and the Holocaust.

Three Prominent Deniers and Denial Groups

One of the prominent HIV denial groups currently is Christine Maggiore's "Alive and Well" (formerly "HEAL," Health Education AIDS Liaison) (http://www.aliveandwell.org/). Maggiore's life story is at the center of this group. Diagnosed with HIV in 1992, Maggiore claims she has since been symptom-free for the past 14 years without the use of antiretroviral drugs, including protease inhibitors [10]. She has risen to prominence, and been embroiled in controversy, in recent years after giving birth to and openly breast-feeding her two children, Charles and Eliza Jane. She had neither child tested for HIV, and did not take antiretroviral medication

- "A solid understanding of the scientific method may not eliminate science denial, but it may act as a buffer against the further spread of such denialist beliefs."

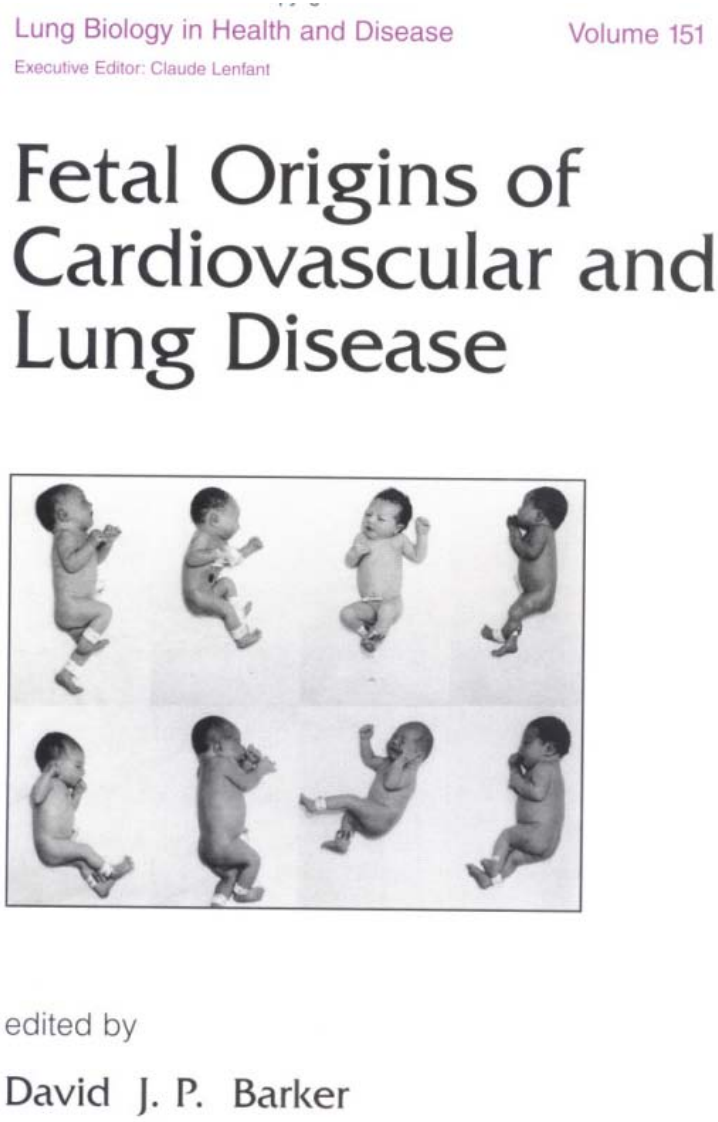
Sufficient and component causes

- 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
  - latency period is not the same as induction period; latency period refers to time delay between disease occurrence and its detection (i.e. diagnosis)
- Component causes can act far apart in time
  - e.g. induction time for adenocarcinoma of vagina in those exposed to DES
  - e.g. fetal origins of adult disease hypothesis

- Blocking the action of any component cause prevents the completion of the sufficient cause and therefore prevents the disease by that pathway [but other pathways remain possible]
- A strong cause is a component cause that plays a causal role in a large proportion of the cases, whereas a weak cause would be a causal component in a small proportion of cases
  - E.g. smoking and lung cancer vs radon gas and lung cancer
  - Strength of a cause necessarily depends on the prevalence of other causal factors that produce the disease

Rothman, 2002

Economist, 2003



A must read paper!

Causation and Causal Inference in Epidemiology

Kenneth J. Rothman, DrPH, Sander Greenland, MA, MS, DrPH, C Stat

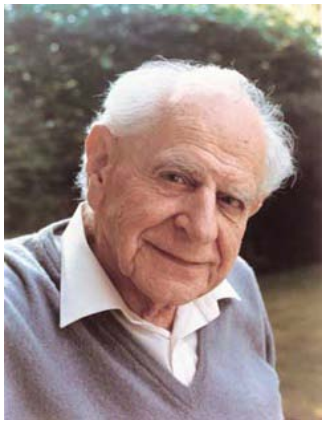
Concepts of cause and causal inference are largely self-taught from early learning experiences. A model of causation that describes causes in terms of sufficient causes and their component causes illuminates important principles such as multicausality, the dependence of the strength of component causes on the prevalence of complementary component causes, and interaction between component causes.

Philosophers agree that causal propositions cannot be proved, and find flaws or practical limitations in all philosophies of causal inference. Hence, the role of logic, belief, and observation in evaluating causal propositions is not settled. Causal inference in epidemiology is better viewed as an exercise in measurement of an effect rather than as a criterion-guided process for deciding whether an effect is present or not. (*Am J Public Health.* 2005;95:S144–S150. doi:10.2105/AJPH.2004.059204)

Causality: Philosophy of scientific inference

- Philosophy of scientific inference
  - Inductivism
    - scientific research proceeds from observations to theories
  - Refutationism or falsificationism
    - science proceeds in the opposite direction, beginning with scientific theories or “conjectures”, and then conducting experiments and eliminating those theories that are falsified by results

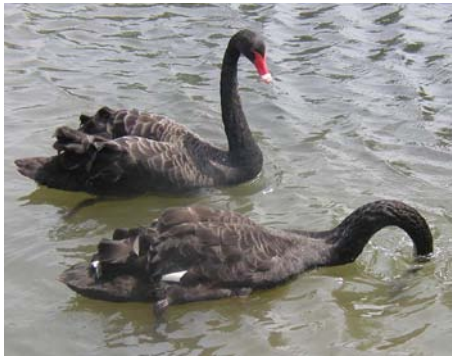
David Hume (1711 – 1776)



Sir Karl Popper (1902 – 1994)

Causality: Philosophy of scientific inference

- Conjecture and refutation
  - Popper proposed falsification as a solution to the problem of induction.
  - Popper noticed that although a singular existential statement such as 'there is a white swan' cannot be used to affirm a universal statement, it can be used to show that one is false:
    - the singular observation of a black swan serves to show that the universal statement 'all swans are white' is false



Wikipedia

Causal inference in epidemiology

- Causal inference in epidemiology
  - Causal criteria
    - E.g. Koch's postulates, Hill's 'criteria'
  - Testing competing theories

Causal criteria: Koch's postulates

- Koch (1884) provided a framework for identifying causes of infectious disease
- Koch's postulates:
  - the agent has to be present in every case of the disease
  - the agent has to be isolated and grown in pure culture
  - the agent has to cause disease when inoculated into a susceptible animal and the agent must then be able to be recovered from that animal and identified

Causal criteria: 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: *Helicobacter pylori* and duodenal ulcer: Evidence suggesting causation. Dig Dis Sci 37:769-772, 1992; and DeCross AJ, Marshall BJ: The role of *Helicobacter pylori* in acid-peptic disease. Am J Med Sci 306:381-391, 1993.

Table 2–2. "Causal criteria" of Hill

Criterion	Problems with the criterion
1. Strength	Strength depends on the prevalence of other causes and, thus, is not a biologic characteristic; could be confounded
2. Consistency	Exceptions are understood best with hindsight
3. Specificity	A cause can have many effects
4. Temporality	It may be difficult to establish the temporal sequence between cause and effect
5. Biologic gradient	Could be confounded; threshold phenomena would not show a progressive relation
6. Plausibility	Too subjective
7. Coherence	How does it differ from consistency or plausibility?
8. Experimental evidence	Not always available
9. Analogy	Analogies abound

Rothman, 2002

Causal inference in epidemiology

- Instead of causal criteria, it may be desirable to put forward multiple theories and test them out systematically
  - Example: toxic shock syndrome: chemical vs infectious theory
- Ken Rothman was asked in an interview, “Which paper has had the most effect on your work?”
  - “Strong Inference” by John R Platt (Science 1964) portrays a systematic method of scientific thinking that can convert pedestrian progress into astonishing success in any branch of science [KJ Rothman]

Strong inference

16 October 1964, Volume 146, Number 3642

SCIENCE

Strong Inference

Certain systematic methods of scientific thinking may produce much more rapid progress than others.

John R. Platt

"nature" or the experimental outcome chooses—to go to the right branch or the left; at the next fork, to go left or right; and so on. There are similar branch points in a "conditional computer program," where the next move depends on the result of the last calculation. And there is a "conditional inductive tree" or "logical tree" of this kind written out in detail in many first-year chemistry books, in the table of steps for qualitative analysis of an unknown sample, where the student is led through a real problem of con-

1) Devising alternative hypotheses;

2) Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;

3) Carrying out the experiment so as to get a clean result;

1′) Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain; and so on.

So, finally, what is epidemiology?

- “Study of the occurrence and distribution of health-related states or events in specified populations, including the study of determinants influencing such states, and the application of this knowledge to control the health problems." [Porta, IEA Dictionary, 2008]

- “Epidemiology is the study of how disease is distributed in populations and the factors that influence or determine this distribution” [Gordis, 2008]

- “Applied, practice-oriented research intended to advance the general knowledge base for community and clinical medicine” [adapted from Miettinen OS, 2007]

- “Application of the scientific method to health research” [adapted from Rothman KJ, 2002]

Going back to the questions posed earlier

- Where is the guarantee that causal claims in the literature are true?
  - No guarantee, at all!
- Could epidemiological studies be wrong?
  - Almost definitely yes!
- Can they make misleading conclusions?
  - Absolutely!
- How can we know when a study result is incorrect?
  - No way of knowing (since we don't know the "truth"), but we can and should be very skeptic of all epidemiologic research (including our own!)
- Is common sense adequate to judge and interpret epidemiologic literature?
  - No, to be skeptical, mere common sense is not adequate
  - As Rothman says, by mastering epidemiology, it is possible to 'educate and refine our common sense'

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

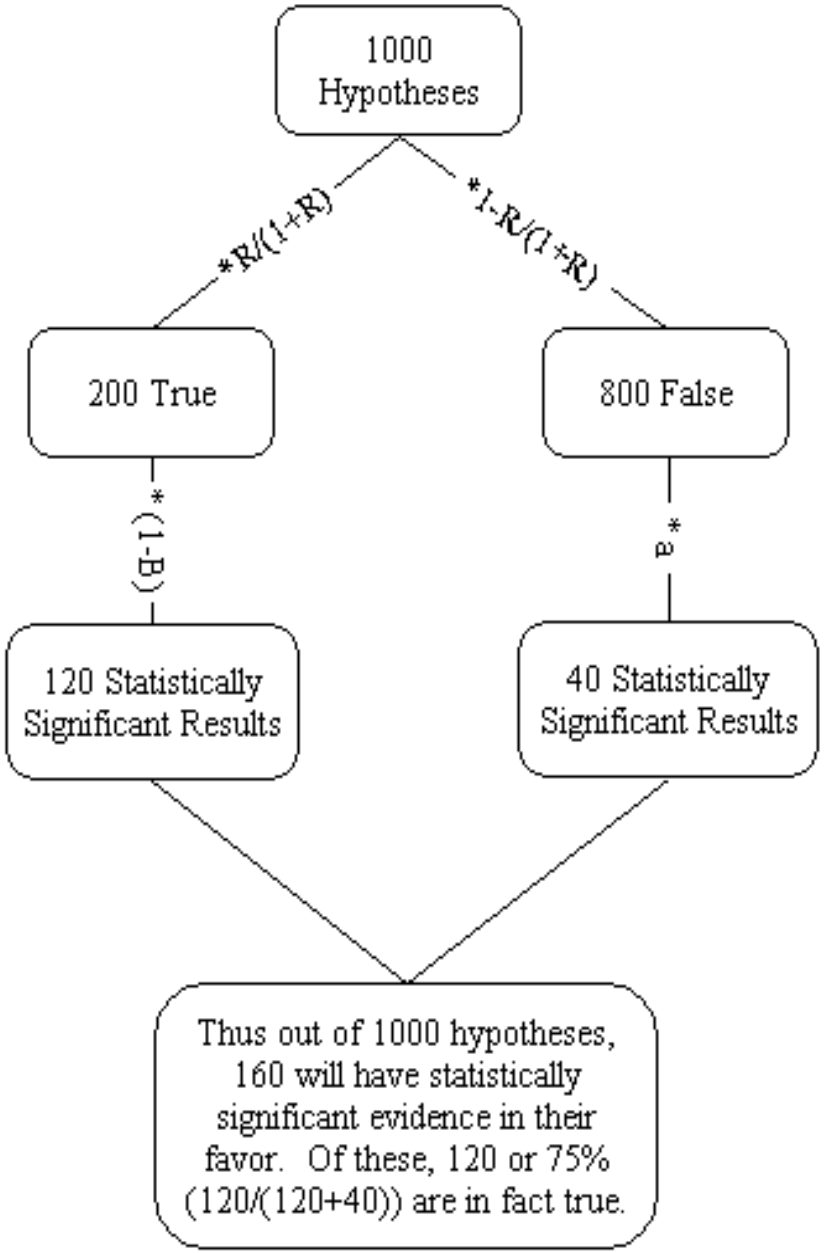
Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2



http://www.marginalrevolution.com/marginalrevolution/2005/09/why_most_publis.html

Importance of replication: Genetic basis for depression?

Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene

Avshalom Caspi,[1,2] Karen Sugden,[1] Terrie E. Moffitt,[1,2]*
Alan Taylor,[1] Ian W. Craig,[1] HonaLee Harrington,[2]
Joseph McClay,[1] Jonathan Mill,[1] Judy Martin,[3]
Antony Braithwaite,[4] Richie Poulton[3]

In a prospective-longitudinal study of a representative birth cohort, we tested why stressful experiences lead to depression in some people but not in others. A functional polymorphism in the promoter region of the serotonin transporter (5-HTT) gene was found to moderate the influence of stressful life events on depression. Individuals with one or two copies of the short allele of the 5-HTT promoter polymorphism exhibited more depressive symptoms, diagnosable depression, and suicidality in relation to stressful life events than individuals homozygous for the long allele. This epidemiological study thus provides evidence of a gene-by-environment interaction, in which an individual's response to environmental insults is moderated by his or her genetic makeup.

18 JULY 2003 VOL 301 SCIENCE www.sciencemag.org

A study, published in 2003 in *Science*, created a sensation among scientists and the public because it offered the first specific, plausible explanation of why some people bounce back after a stressful life event while others plunge into lasting despair.

Genetic basis for depression?

Interaction Between the Serotonin Transporter Gene (*5-HTTLPR*), Stressful Life Events, and Risk of Depression

A Meta-analysis

Neil Risch, PhD
Richard Herrell, PhD
Thomas Lehner, PhD
Kung-Yee Liang, PhD
Lindon Eaves, PhD
Josephine Hoh, PhD
Andrea Griem, BS
Maria Kovacs, PhD
Jurg Ott, PhD
Kathleen Ries Merikangas, PhD

Context Substantial resources are being devoted to identify candidate genes for complex mental and behavioral disorders through inclusion of environmental exposures following the report of an interaction between the serotonin transporter linked polymorphic region (*5-HTTLPR*) and stressful life events on an increased risk of major depression.

Objective To conduct a meta-analysis of the interaction between the serotonin transporter gene and stressful life events on depression using both published data and individual-level original data.

Data Sources Search of PubMed, EMBASE, and PsycINFO databases through March 2009 yielded 26 studies of which 14 met criteria for the meta-analysis.

Study Selection Criteria for studies for the meta-analyses included published data on the association between *5-HTTLPR* genotype (SS, SL, or LL), number of stressful life events (0, 1, 2, ≥3) or equivalent, and a categorical measure of depression defined by the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) or the *International Statistical Classification of Diseases, 10th Revision (ICD-10)* or use of a cut point to define depression from standardized rating scales. To maximize our ability to use a common framework for variable definition, we also requested original data from all studies published prior to 2008 that met inclusion criteria. Of the 14 studies included in the meta-analysis, 10 were also included in a second sex-specific meta-analysis of original individual-level data.

Data Extraction Logistic regression was used to estimate the effects of the number of short alleles at *5-HTTLPR*, the number of stressful life events, and their interaction on depression. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated separately for each study and then weighted averages of the individual estimates were obtained using random-effects meta-analysis. Both sex-combined and sex-specific meta-analyses were conducted. Of a total of 14 250 participants, 1769 were classified as having depression; 12 481 as not having depression.

Results In the meta-analysis of published data, the number of stressful life events was significantly associated with depression (OR, 1.41; 95% CI,1.25-1.57). No association was found between *5-HTTLPR* genotype and depression in any of the individual studies nor in the weighted average (OR, 1.05; 95% CI, 0.98-1.13) and no interaction effect between genotype and stressful life events on depression was observed (OR, 1.01; 95% CI, 0.94-1.10). Comparable results were found in the sex-specific meta-analysis of individual-level data.

Conclusion This meta-analysis yielded no evidence that the serotonin transporter genotype alone or in interaction with stressful life events is associated with an elevated risk of depression in men alone, women alone, or in both sexes combined.

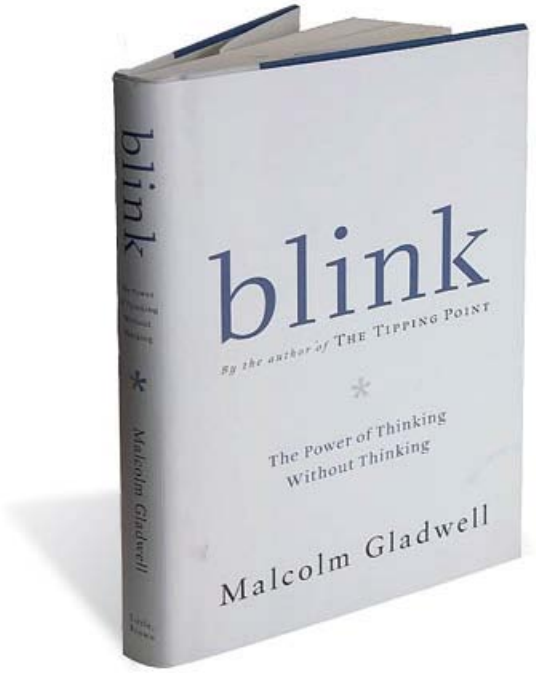
JAMA. 2009;301(23):2462-2471 www.jama.com

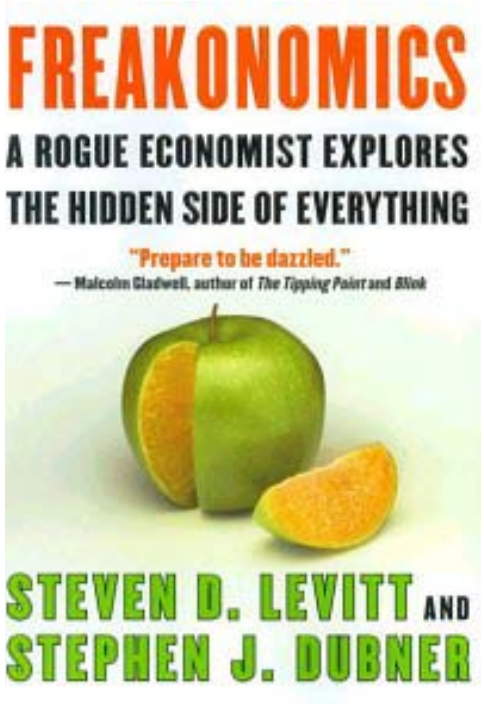
THE SUCCESSFUL STATISTICAL identification and independent replication of numerous genetic markers in association studies have confirmed the utility of the genome-wide approach for the detection of genetic markers for complex disorders.[1,2] However, recent genome-wide association studies have also indicated that most common genetic risks, at least when studied individually, are modest in magnitude, with relative risks in the range of 1.3 or less.[2] This suggests that complex disorders result from the combination of numerous individual genetic and environmental contributors, with the potential for interactions among them. However, there is a lack of consensus regarding whether gene × gene or gene × environment interactions should be examined at the stage of gene de-

Meta-analysis, published in JAMA in 2009, found no evidence that the serotonin transporter genotype alone or in interaction with stressful life events is associated with an elevated risk of depression in men alone, women alone, or in both sexes combined.

"it is critical that health practitioners and scientists in other disciplines recognize the importance of replication of such findings before they can serve as valid indicators of disease risk"

Two approaches to making decisions: which reflects epidemiology best?

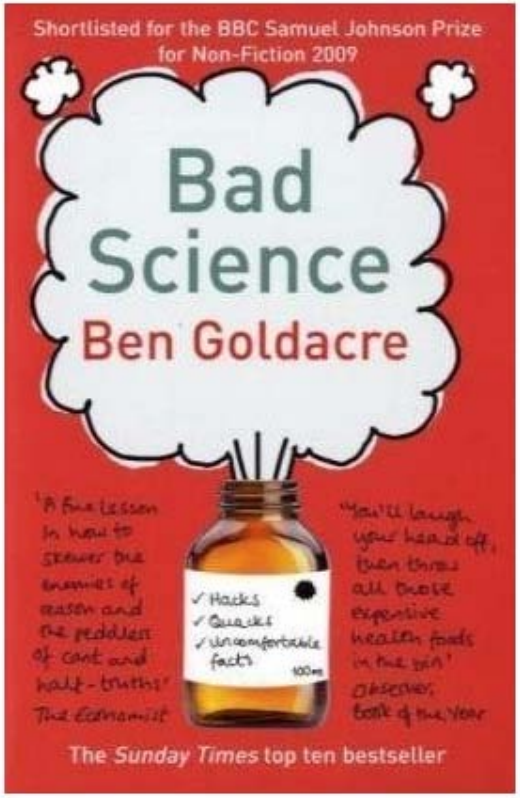


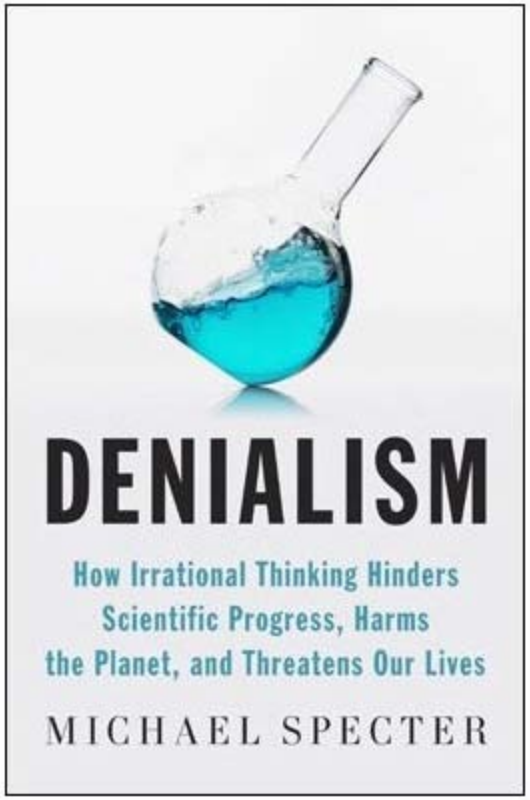


"it's about rapid cognition, about the kind of thinking that happens in a blink of an eye.."

"it's all about using information about the world around us to get to the heart of what's really happening under the surface..."

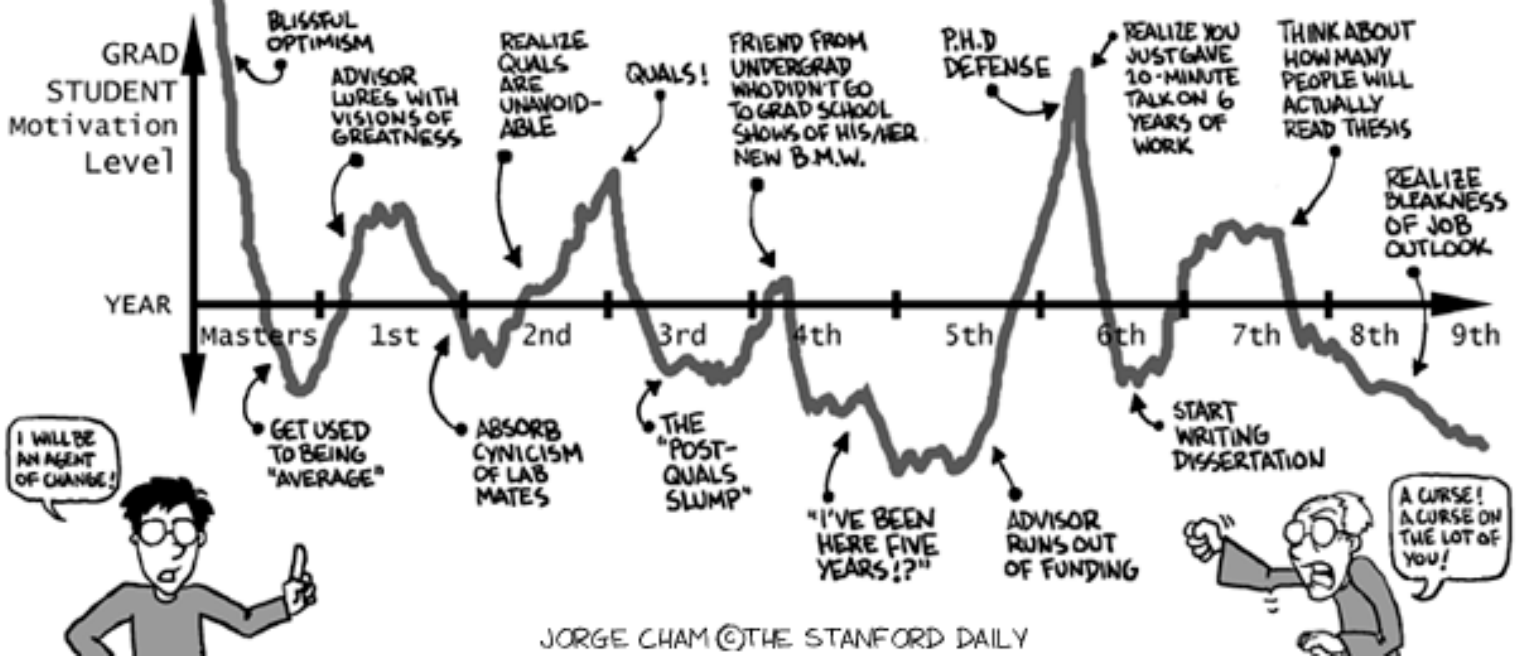
Two books worth reading!

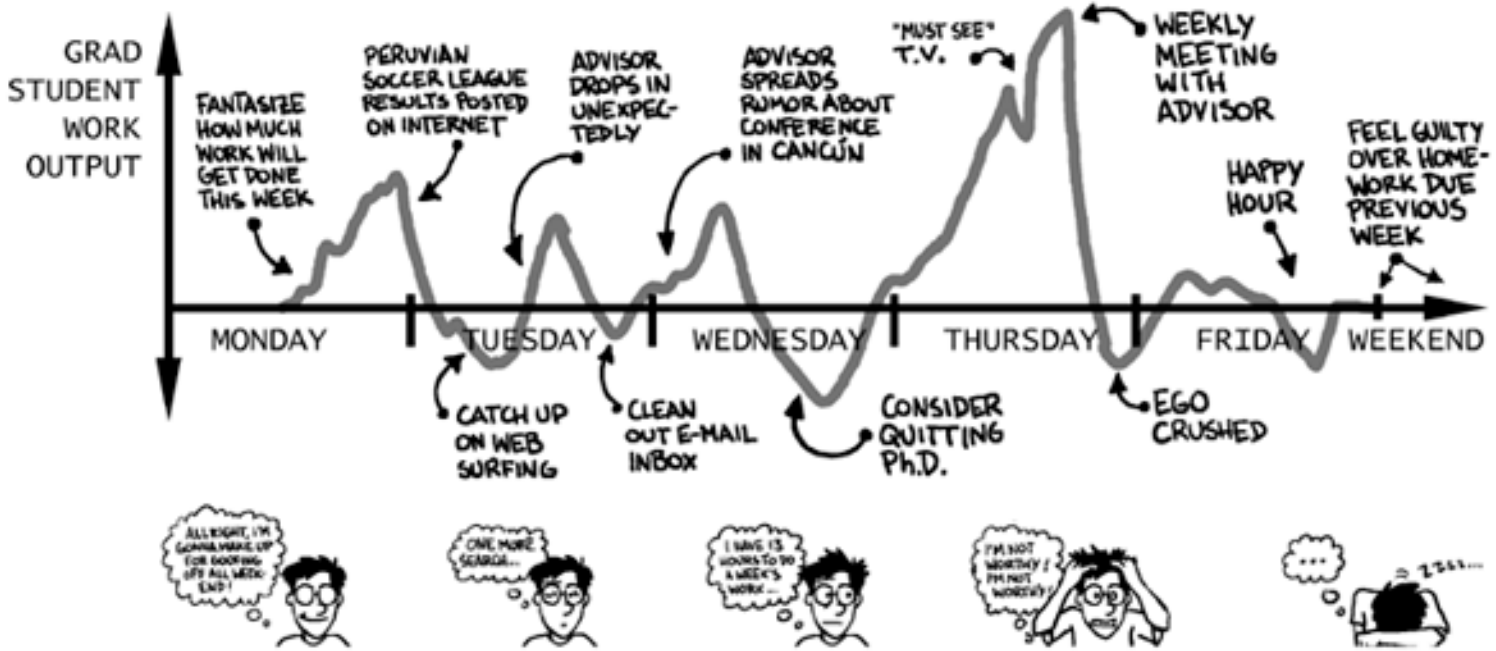




Readings

- Article:
  - Rothman & Greenland. Causation and causal inference in epidemiology. Am J Pub Health 2005.
- Rothman text:
  - Chapters 2: What is causation?
- For ‘extra credit’:
  - Harry Potter and the Prisoner of Azkaban
  - Matrix Trilogy
  - Freakonomics, Steven Levitt & Stephen Dubner
  - Blink, Malcolm Gladwell


GRAD STUDENT Motivation Level
YEAR
Masters
1st
2nd
3rd
4th
5th
6th
7th
8th
9th
BLISSFUL OPTIMISM
ADVISOR LURES WITH VISIONS OF GREATNESS
REALIZE QUALS ARE UNAVOID-ABLE
QUALS!
FRIEND FROM UNDERGRAD WHO DIDN'T GO TO GRAD SCHOOL SHOWS OF HIS/HER NEW B.M.W.
P.H.D DEFENSE
REALIZE YOU JUST GAVE 20-MINUTE TALK ON 6 YEARS OF WORK
THINK ABOUT HOW MANY PEOPLE WILL ACTUALLY READ THESIS
REALIZE BLEAKNESS OF JOB OUTLOOK
I WILL BE AN AGENT OF CHANGE!
GET USED TO BEING "AVERAGE"
ABSORB CYNICISM OF LAB MATES
THE "POST-QUALS SLUMP"
"I'VE BEEN HERE FIVE YEARS!?"
ADVISOR RUNS OUT OF FUNDING
START WRITING DISSERTATION
A CURSE! A CURSE ON THE LOT OF YOU!
JORGE CHAM ©THE STANFORD DAILY


GRAD STUDENT WORK OUTPUT
MONDAY
TUESDAY
WEDNESDAY
THURSDAY
FRIDAY
WEEKEND
FANTASIZE HOW MUCH WORK WILL GET DONE THIS WEEK
PERUVIAN SOCCER LEAGUE RESULTS POSTED ON INTERNET
ADVISOR DROPS IN UNEXPEC-TEDLY
ADVISOR SPREADS RUMOR ABOUT CONFERENCE IN CANCÚN
"MUST SEE" T.V.
WEEKLY MEETING WITH ADVISOR
HAPPY HOUR
FEEL GUILTY OVER HOME-WORK DUE PREVIOUS WEEK
CATCH UP ON WEB SURFING
CLEAN OUT E-MAIL INBOX
CONSIDER QUITTING Ph.D.
EGO CRUSHED
ALL RIGHT, I'M GONNA MAKE UP FOR GOOFING OFF ALL WEEK! END!
ONE MORE SEARCH...
I HAVE 15 HOURS TO DO A WEEK'S WORK...
I'M NOT WORTHY! I'M NOT WORTHY!
ZZZZ...
JORGE CHAM ©THE STANFORD DAILY