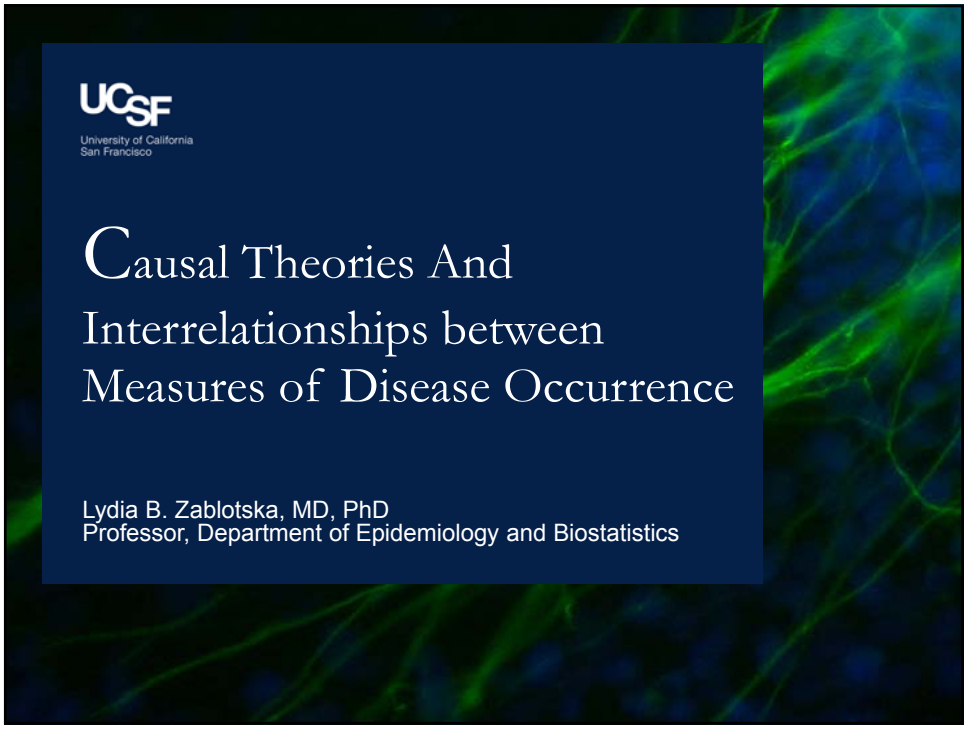


Learning Objectives

- Discuss how causal inference is central to the role of epidemiology
- Provide a brief overview of the history of causal thinking through the years
- Introduce theories of causal inference
- Describe causal models and heuristics
  - i. Causal guidelines (Hill)
  - ii. Sufficient-component cause model (Rothman)
    - Describe (and critique) Rothman's causal heuristic
  - iii. Counterfactual model (Greenland and Morgenstern)
    - Counterfactual effect measures: rate ratios, risk ratios and odds ratios
    - Effect measures vs. measures of association
    - Measures of attributable risk
  - iv. Graphical models (e.g., population pyramids, causal diagrams, DAGs)
  - v. Complex systems approach (e.g., agent-based modeling and g-estimation)
- Discuss how epidemiologic thinking leads to causal inference

2 UCSF

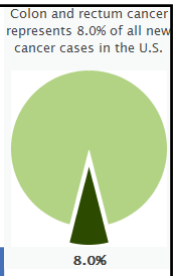
Organization of the Course

- See Course Syllabus and Welcome video on the CLE

UCSF

Example: "The Fiber Yarn"

Study of the association between fiber intake and risk of colorectal cancer



Incidence rates of colorectal cancer per year in the U.S.			
	SEER 2000	SEER 2010	SEER 2013
Males	64 per 100,000	47 per 100,000	44 per 100,000
Females	47 per 100,000	36 per 100,000	34 per 100,000

- 4.4% of men and women born today will be diagnosed with cancer of the colon and rectum at some time during their lifetime
- 4th leading cause of cancer (after prostate, breast and lung)
- 2nd leading cause of death from cancer (after lung)

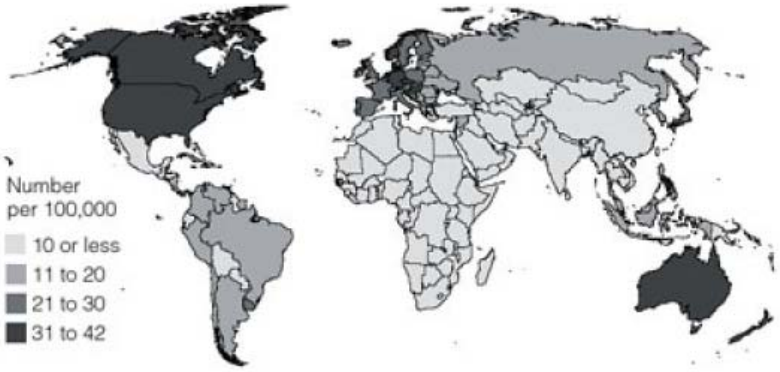
 Howlader et al. 2016

UCSF

UCSF

What Seemed to Be a Correlation . . .

It appeared that colorectal cancer might be related to fiber intake, because it occurs more often in countries with Western diets, which tend to be lower in fiber. The map shows the incidence of colorectal cancer in women for each country.



5 5

UCSF

… Was Researched and Tested… a.k.a. "Untangling *the fiber yarn*"

Sebastian and Mostoslavsky, *Cancer Discovery* 2014

- Lin J et al. Cancer Causes Control. 2005 - **Legume fiber** and/ or other related sources may reduce risk of colorectal cancer. "
- Shin A et al. Int J Cancer. 2006 - **Calcium** may be protective against colorectal cancer development …"
- A. Schatzkin et al. Am J Clin Nutr. 2007 - **Whole-grain intake** was inversely associated with colorectal cancer risk...
- K. Uchuda et al. Scand J Gastroenterol. 2010 - **Rice consumption** was associated with a decreased risk of colorectal cancer
- D. Aune et al. BMJ. 2011 - **Cereal fibre and whole grains**, was associated with a reduced risk of colorectal cancer

6

UCSF

… Saga Continues …
a.k.a. "Untangling *the fiber yarn*"

Sebastian and Mostoslavsky, *Cancer Discovery* 2014

- Kantor ED et al. Cancer Epidemiol Biomarkers Prev. 2014 – The associations between **genetic susceptibility loci** and colorectal cancer are not strongly modified by fiber consumption.
- Donohoe et al. Cancer Discover. 2014 - Fiber does have a potent tumor-suppressive effect but in a **microbiota- and butyrate-dependent manner**.

2015: … Back to square one???

- **Kunzmann AT** et al. **Am J Clin Nutr.** 2015 – "…Individuals consuming the highest intakes of dietary fiber, **particularly from cereals and fruit**, have reduced risks of incident colorectal adenoma."

7

UCSF

Risk Factor Epidemiology

The New York Times

Saturday, September 26, 2015 | Today

Why Everything Is Bad for You

Article after article seems eager to convince us that food is killing us. Might as well eat bologna. Because you just can't win.

"Week after week of cause after cause."

Kelsey J, *SER* 1983

8

UCSF

We Are Not Doing the General Public Any Favors With Our Research…


The New York Times

BOOKS

Book Review: 'Ending Medical Reversal' Laments Flip-Flopping

BY ABIGAIL ZUGER, M.D.

Ever-changing advice on treatment and care is demoralizing both doctors and patients, the authors claim.

UCSF

Making the Case For…

Moving from Risk Factor Epidemiology to Modern Epidemiology

"The present era of epidemiology is coming to a close. The focus on risk factors at the individual level -- the hallmark of this era -- will no longer serve. We need to be concerned equally with causal pathways at the societal level and with pathogenesis and causality at the molecular level."

Susser E, *Choosing a Future for Epidemiology*, AJPH 1996

UCSF

UCSF

Classical Definition of Epidemiology

"… the study of the distribution and determinants of disease frequency in human populations"

MacMahon and Pugh, **Epidemiology: Principles and methods** (1970)

UCSF

Classical Definition of Epidemiology … Updated

"… the study of the distribution and determinants of disease frequency in human populations"

We also add:

… AND the application of this study to

- ***control*** **health problems**
- ***improve*** **public health**

UCSF

UCSF

Modern Definition of Epidemiology

"…science that focuses on the occurrence of disease in its broadest sense, with the fundamental aim to *understand and to control its causes*"

Rothman and Greenland *in* **Handbook of Epidemiology** *by* Ahrens and Pigeot (2007)

- i.e., rather than on the natural history or some other aspect of the disease

13

UCSF

Why worry about causes of diseases?

- So that we can intervene
- So that we can reduce or prevent disease

Causal Inference: a process of determining causal and preventive factors

➢**Question:** How could we determine causes of diseases?

14

UCSF

What is a cause?

"A cause is something that makes a difference. Insofar as epidemiology is a science...[that] aims to discover the cause of health states, the search includes all determinants of health outcomes. These may be both active agents... and static conditions such as the attributes of persons and places."

Susser M, *Am J Epidemiol* 1991

15

UCSF

Epidemiology defined:

- Aims to find causes of diseases and to explain varying patterns of disease occurrence across populations and groups
- The basic science or one of the pillars of public health
- Way of thinking and logically structuring scientific inquiry in public health
- Scientific discipline with roots in biology, medicine, logic, and the philosophy of science

16

UCSF

Societal origins of epidemiology

- Epidemiology affects the daily lives of most people
- Comes from the Greek words *epi*, *demos*, and *logos* meaning '*the study of people*'
- Originated in the Sanitary Era (XIX century) out of necessity to improve the economic productivity by decreasing squalor of the industrial slums
- Epidemiology is the result of the evolution of progressive thinking and our understanding of the basic human rights

17 UCSF

And since it is the purpose of epidemiology to…

- Identify factors that cause the distribution of disease

18 UCSF

And since it is the purpose of epidemiology to…

- Identify factors that cause the distribution of disease
- This must be the most important lecture of the course…

UCSF

Brief history of causal thinking through the years

"If you don't know history, then you don't know anything. You are a leaf that doesn't know it is part of a tree. "

Michael Crichton, Timeline (1999)

UCSF

UCSF

Historical developments in the understanding of causes of diseases

1. Sanitary era

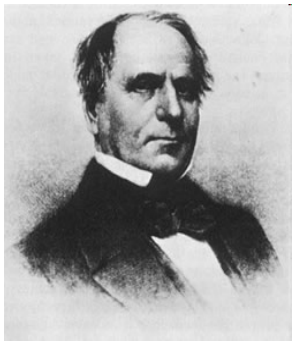
➢paradigm: miasma

Miasma theory of Sydenham:

- foul emanations from soil, water and air cause all diseases
- poverty is at the core of all ills, it is a cause rather then a consequence of disease

The Public Health Act of 1848

- Decaying organic matter →insanitary conditions → foul emanations →diseases →poverty → high birth rates among poor



Edwin Chadwick (1800 – 1890)

21

UCSF

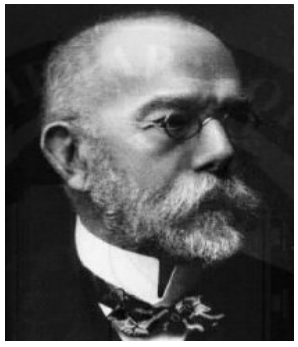
Historical developments in the understanding of causes of diseases

2. Infectious disease era

➢paradigm: germ theory

Discovery of causal agents of anthrax, tuberculosis and cholera by R. Koch

- Bacillus anthracis (1877)
- Mycobacterium tuberculosis (1882)
- Vibrio cholerae (1883)



Robert Koch (1843 – 1910)

22

UCSF

Causal Inference: Henle-Koch postulates for causation

1890:

- The organism is always found with the disease
- The organism is not found with any other disease
- The organism, isolated from one who has the disease, and cultured through several generates, produces the disease (in experimental animals)

Jacob Henle (1809 – 1885)

23

UCSF

Historical developments in the understanding of causes of diseases

3. Risk factor epidemiology or chronic disease era

➢paradigm: black box

"Web of causation"

MacMahon et al., Epidemiologic methods (1960)

- All factors are at the same level
- Diseases can be prevented by cutting a few strands of the web
- Does not elucidate societal forces or their relation to health

"… too much statistics takes away all the pleasure and the message of epidemiology."

Brian MacMahon (1923 – 2007)

24

UCSF

Historical developments in the understanding of causes of diseases

4. Ecoepidemiology

➢paradigm: Chinese (nesting) boxes

- Conceptual approach combining molecular, societal, and population-based aspects to study a health-related problem.
- People are not only individuals but also members of communities (social context).
- Helps to recognize broad dynamic patterns and disease in its social context.
- Places exposure, outcome and risk in societal context.

Mervyn Susser (1921 – 2014)

25 Susser, ***Eras in Epidemiology: The Evolution of Ideas*** (2009) UCSF

Where are we now?

5? From Risk Factor to Consequentialist Epidemiology

- We have strayed away from the "parsimonious, clear and useful" definition of epidemiology by abandoning our roots in the Sanitary era epidemiology and moving away from trying to improve population health.
- 85% of all papers published in the four leading epi journals in 2012 were concerned with etiology and only 15% with public health interventions.
- Epidemiology needs to link with Clinical Implementation Science and Translational Research
- **The main focus should be on developing new approaches and methods to study etiology:**
  - **Traditional questions "Who, What, When, Where, and Why?" of Descriptive Epidemiology**
  - ➢ **Add "So What?" and "How Much?"**
  - ➢ **Add "What Matters Most" lens**

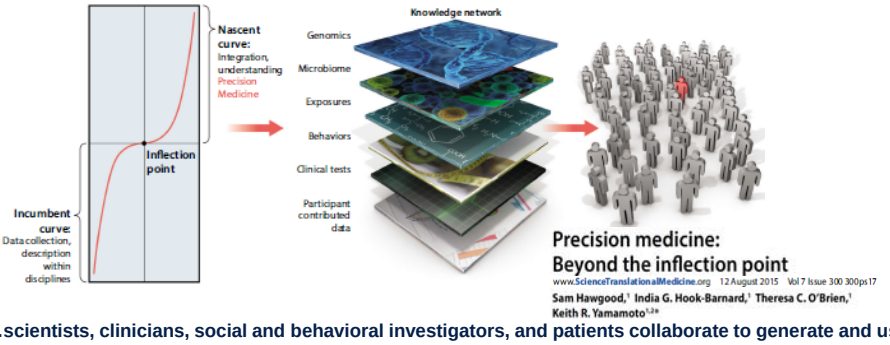
Sandro Galea

Foege *APHA* 1983
Cates *SER* 1994
Galea *Am J Epidemiol* 2013
Keyes & Galea *Ann Epid* 2015

26 UCSF

Where are we now?

6? Precision Medicine Era



"…scientists, clinicians, social and behavioral investigators, and patients collaborate to generate and use massive data networks that access, aggregate, integrate, and analyze information from huge patient cohorts, healthy populations, and experimental organisms in order to determine mechanisms of normal and disease processes and provide precise health advice, diagnoses, and treatments for each individual."

- **Questions:** Where does Epidemiology fit in the Precision Medicine Era? How could we advance precision medicine agenda and make sure that new treatments benefit all groups of our society???

27

UCSF

Causal inference

Goal of epidemiology:

learn causes of diseases and factors that could prevent or delay disease development

Causal inference:

a process of determining causal and preventive factors

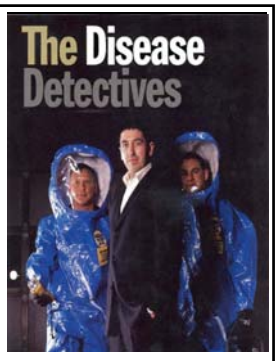
28

UCSF

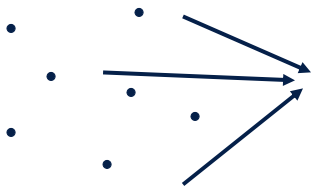
Theories of causal inference

- Deductive reasoning
- Inductivism
- Refutationism
- Bayesianism ➤ Epi 265

Deductive reasoning



- George Simenon's *Inspector Maigret*
- Arthur Conan Doyle's *Sherlock Holmes*
- Agatha Christie's *Hercule Poirot*
- Modern disease detectives: *Sandro Galea*



Pull the clues together, arrive at generalization, i.e. *deduct* the answer

Inductive reasoning

Conditional Inductive Tree:

formulate laws based on limited observations of recurring phenomenal patterns:

Specification of alternative hypotheses

↓

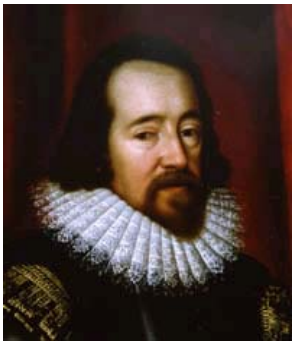
Design of crucial experiments to test these hypotheses

↓

Exclusion of some alternatives

↓

Adoption of what is left (for the time being)



Sir Francis Bacon (1561 – 1626)

Bacon, ***Novum Organum Scientiarum*** *(1620)*

31

UCSF

David Leonhardt, July 2, 2015, NY Times

- CONFIRMATION bias - Not only are people more likely to believe information that fits their pre-existing beliefs, but they're also more likely to go looking for such information (Cathcart Wason 1960) → the internal yes-man (and yes-woman) tendency → suggests that we are much more likely to think about why something might go right than wrong.
- Often, people never even think about asking questions that would produce a negative answer when trying to solve a problem
- When you want to test a theory, don't just look for examples that prove it. When you're considering a plan, think in detail about how it might go wrong.
- When you seek to disprove your idea, you sometimes end up proving it — and other times you can save yourself from making a big mistake. And even if you think that you are right, you need to make sure you're asking questions that might actually produce an answer of No.

32

UCSF

A Quick Puzzle to Test Your Problem Solving → Are you a Yes-Man or Yes-Woman?

http://www.nytimes.com/interactive/2015/07/03/upshot/a-quick-puzzle-to-test-your-problem-solving.html

A Quick Puzzle to Test Your Problem Solving

By DAVID LEONHARDT and YOU JULY 2, 2015

A short game sheds light on government policy, corporate America and why no one likes to be wrong.

Here's how it works:

We've chosen a rule that some sequences of three numbers obey — and some do not. Your job is to guess what the rule is.

We'll start by telling you that the sequence 2, 4, 8 obeys the rule:

2	4	8	Obeys the rule

Now it's your turn. Enter a number sequence in the boxes below, and we'll tell you whether it satisfies the rule or not. You can test as many sequences as you want.

Your guesses:

16	32	64	Yes!

When you think you know the rule, describe it in words below and then submit your answer. **Make sure you're right; you won't get a second chance.**

UCSF

Deductive vs. inductive reasoning

Deductive reasoning applies general principles to reach specific conclusions, whereas inductive reasoning examines specific information, perhaps many pieces of specific information, to derive a general principle.

UCSF

UCSF

Criticism of inductive reasoning

- D. Hume suggested that because inductive reasoning relies on inferences made from empirical observations, thy could be erroneous because observations depend on sense perceptions
- K. Popper developed a theory of *refutationism,* which proposes to subject all new hypotheses to rigorous tests aimed at falsifying them in preference to repetitions of the original observations that add little beyond the weak corroboration of the initial findings

UCSF

Causal models and heuristics

i. Causal guidelines (Hill)
ii. Sufficient-component cause model (Rothman)
iii. Counterfactual model (Greenland and Morgenstern)
iv. Graphical models (e.g., population pyramids, causal diagrams, DAGs)
v. Complex systems approach (e.g., agent-based modeling and g-estimation)

UCSF

UCSF

Criteria for Causal Inference

1. Strength
2. Consistency
3. Specificity
4. Temporality
5. Biological gradient
6. Plausibility
7. Coherence
8. Experiment
9. Analogy

A. Bradford Hill (1897 – 1991)

37 Hill A.B. *Proc Roy Soc Med 1965*

UCSF

1. Strength of association

- Strong associations are less likely to be caused by chance or bias
- A strong association means a very high or very low relative risk

➢ CAVEAT

Environmental associations with very low relative risks

38

UCSF

2. Consistency

- Replication of findings in different populations under different circumstances, in different times, with different study designs

➢ CAVEATS

- **Lack of consistency does not rule out a causal association, because some effects are produced by their causes only under unusual circumstances**
- **Publication bias**
- **Contradictory findings across different studies are not unusual in studies of weak effects**

39 UCSF

3. Specificity of the association

- Specific exposure associated with only one disease
  - Arises from old Henle-Koch postulates for causation
- Effect has one cause, not multiple causes

➢ CAVEATS

- **Many exposures are linked to multiple diseases**
- **Many diseases have multiple causes**

40 UCSF

4. Temporality

- Exposure must precede disease (cause must precede effect)
- Levels of evidence
- Cross-sectional studies (exposure and disease measured at the same time)
  - e.g., NHANES (National Health and Nutrition Examination Survey) looking at the link between obesity and coronary artery disease
- Case-control studies (compare exposures and risk factors among people with and without the disease)
  - e.g., case-control study investigating the link between radiation exposure among Chornobyl clean-up workers and leukemia
- Cohort studies (follow-up exposed and unexposed to see who will develop the disease)
  - e.g., cohort study of children with thyroid activity measurements taken within 6 weeks after the Chornobyl accident and development of thyroid cancer 15-22 years later)
  - In disease with long latency periods, exposures must precede latency period
  - In chronic diseases, often need long-term exposure for disease induction

41 UCSF

5. Biologic gradient
(dose-response relationship)

- Presence of a dose-response or exposure-response curve with an expected shape
- Changes in exposure are related to trend in risk of disease
- Strong evidence for causal relation suggesting biologic relation

42 UCSF

Ukraine, prevalence

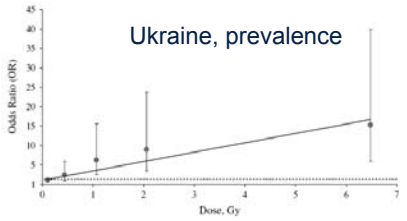
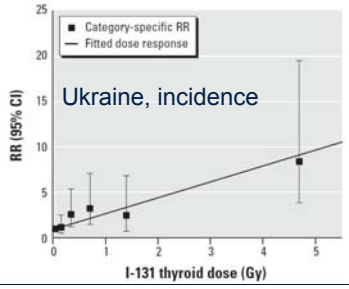


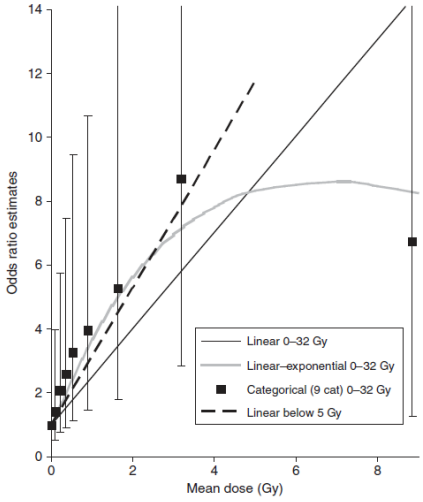
Fig. 2. Odds ratios (ORs) with 95% confidence intervals of thyroid cancer by mean dose for each of five dose categories. The data from Table 2 were plotted, and a fitted dose–response line was constructed using the least squares method fitted to the categorical point estimates. **Dots** = odds ratios; **vertical lines** = 95% confidence intervals; **heavy solid line** = fitted dose–response line; **dashed line** indicates OR = 1.0.

Tronko et al. JNCI 2006



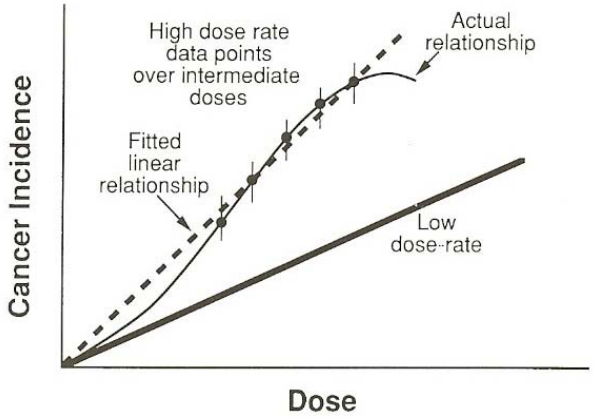
Ukraine, incidence

Belarus, prevalence



43 Brenner et al. 2011 Zablotska et al. 2011 UCSF

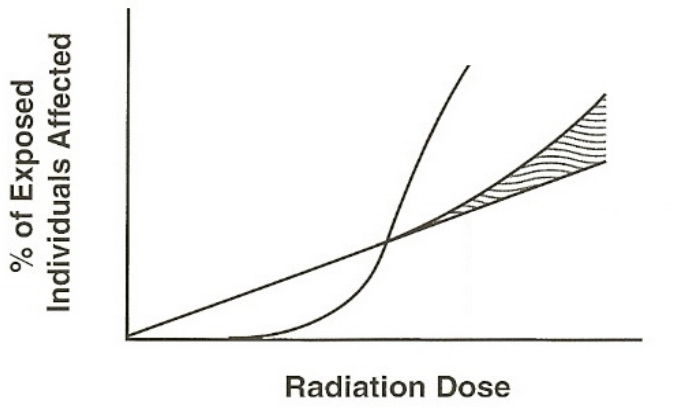
Study of the survivors of atomic bombings in Hiroshima



Association between radiation dose received in 1945 and risk of developing cancer later in life.

44 E. Hall, **Radiobiology for the radiologist** (2000) UCSF

Threshold effect?



45 E. Hall, **Radiobiology for the radiologist** (2000)

UCSF

5. Biologic gradient (dose-response relationship)

- Presence of a dose-response or exposure-response curve with an expected shape
- Changes in exposure are related to trend in risk of disease
- Strong evidence for causal relation suggesting biologic relation

➢ CAVEAT

Thresholds, i.e., no disease past a certain level of exposure

46

UCSF

6. Plausibility

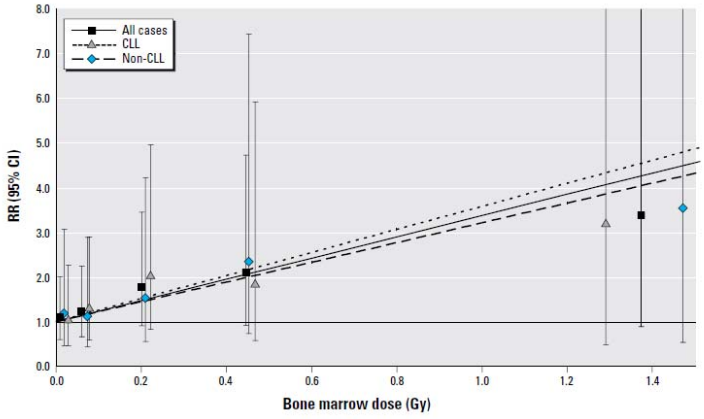
- The proposed mechanism should be biologically (etiologically) plausible
- Reference to a "coherent" body of knowledge

CAVEATS

- New diseases and new causes
- Theoretical plausibility

UCSF

Radiation-associated risks of chronic lymphocytic leukemia vs. other types of leukemia



Zablotska et al. *EHP* 2012.

UCSF

UCSF

7. Coherence with established "facts"

- A cause-and-effect interpretation for an association does not conflict with what is known of the natural history and biology of disease
- Implications:
  - **If a relation is causal, would expect observed findings to be consistent with other data**
  - **Hypothesized causal relations need to be consistent with epidemiologic and biologic knowledge**

➢ CAVEATS

- **Data may not be available yet to directly support proposed mechanism**
- **Science must be prepared to reinterpret existing understanding of disease process in the face of new evidence**

49 UCSF

8. Experiment

- In Hill's article, refers to 'cessation of exposure', i.e. elimination of putative harmful exposure results in the decrease of the frequency of disease

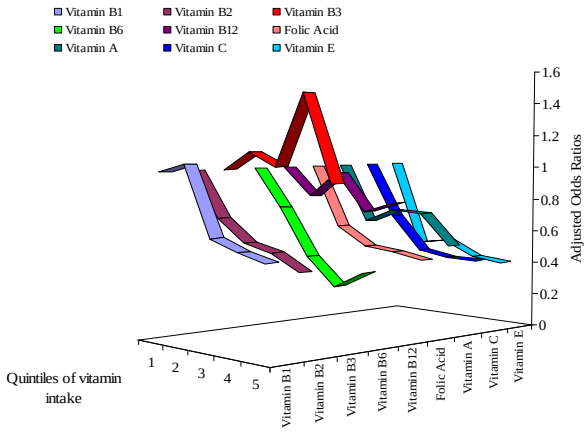
➢ CAVEATS

- **If the pathogenic process has already started, removal of cause does not reduce disease risk**
- **Reduction in disease frequency might not be for etiologic reason hypothesized**

50 UCSF

9. Analogy

Figure 1. Odds Ratios for Quintiles of Vitamin Intake From the Categorical Analysis.



51 Zablotska et al. *EHP* 2008

UCSF

9. Analogy

- Similar exposures can cause similar effects, e.g., medications and infectious agents may cause other birth defects

➢ CAVEAT

- **Limited by the current knowledge**

52

UCSF

Overall caveats to "criteria"

"None of my ... [criteria] can bring undisputable evidence for or against the cause-and-effect hypothesis and none can be required as a sine qua non."

Temporality?

 Hill A.B. *Proc Roy Soc Med 1965*

UCSF

Summary: When is an association causal?

Theory	Smoking is a carcinogen
Hypothesis	Smoking causes lung cancer
Strategy to test the hypothesis	Prospective cohort study
Design, conduct, and analysis of study	Recruit 10,000 doctors, follow for 10 years
Interpretation of results	High RR of lung cancer in smokers

UCSF

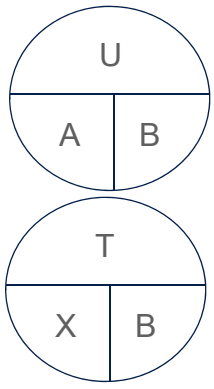
Application of Hill's Guidelines for Causal Inference to the study of the association between fiber intake and risk of colorectal cancer

1. Strength - Yes
2. Consistency - Questionable
3. Specificity - No
4. Temporality - Yes
5. Biological gradient - Yes
6. Plausibility - Yes
7. Coherence - Possible
8. Experiment - Yes
9. Analogy - Yes

55

The causal pie model (sufficient component model)

- Each pie is a theoretical causal mechanism for a given disease ("sufficient cause")
- Each individual instance of disease occurs through a single sufficient cause



56 Rothman, **Epidemiology: An introduction** (2012), Ch 3

What is a cause?

- A cause of a specific disease event [is] an antecedent event, condition or characteristic that was **necessary** for the disease at the moment it occurred, given that other conditions are fixed.
- A cause of a disease is an event, condition, or characteristic that preceded the disease event and **without which the disease event would not have occurred** at all or would not have occurred until some later time.

Kenneth Rothman

57

UCSF

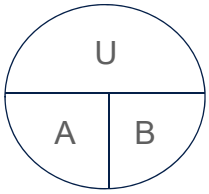
Types of causal relationships

- If a relationships is indeed causal, then…
- Necessary and sufficient
  - E.g., rabies, HIV exposure in AIDS
- Necessary but not sufficient
  - Multiple factors acting in a specific temporal sequence
  - E.g., multistage carcinogenesis
- Sufficient but not necessary
  - E.g., both ionizing radiation and benzene exposure cause leukemia independently
- Neither sufficient nor necessary
  - Many different pathways of getting the same disease

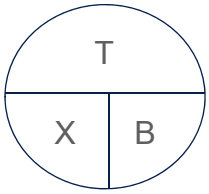
58

UCSF

Sufficient and component causes



Sufficient Cause 1



Sufficient Cause 2

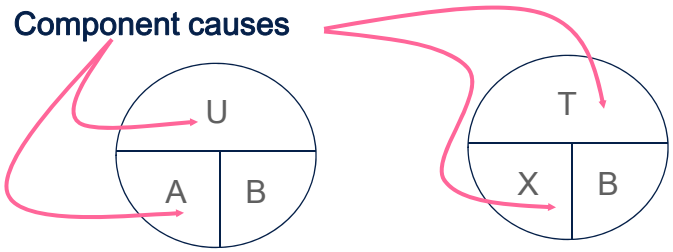
A sufficient cause is a set of minimal conditions or events that inevitably produce disease

59

UCSF

Sufficient and component causes

Component causes



Sufficient Cause 1

Sufficient Cause 2

A sufficient cause is a set of minimal conditions or events that inevitably produce disease

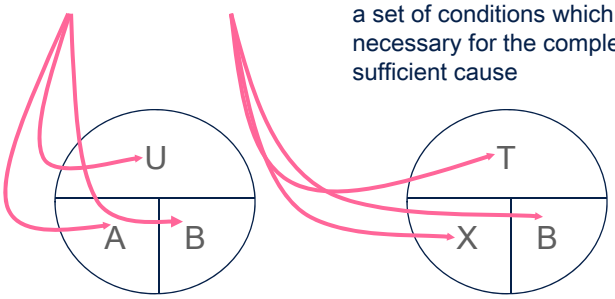
60

UCSF

Sufficient and component causes

Component causes

A **component cause** is any one of a set of conditions which are necessary for the completion of a sufficient cause



Sufficient Cause 1

Sufficient Cause 2

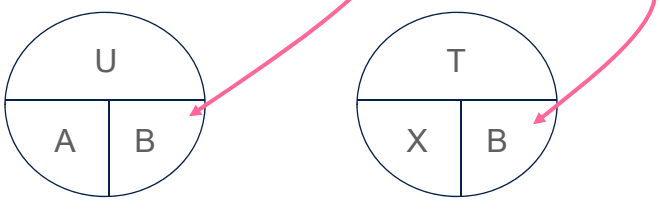
A sufficient cause is a set of minimal conditions or events that inevitably produce disease

61

UCSF

Sufficient and component causes

A **necessary component cause** is a component cause that is a member of every sufficient cause



Sufficient Cause 1

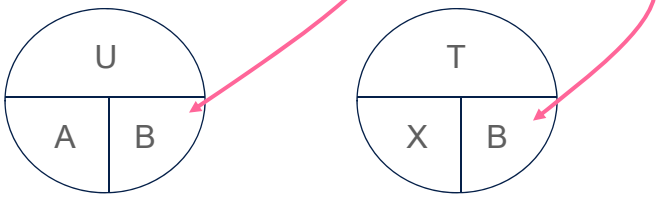
Sufficient Cause 2

62

UCSF

Sufficient and component causes

A strong component cause is a component cause that plays a causal role in a high proportion of cases in the population



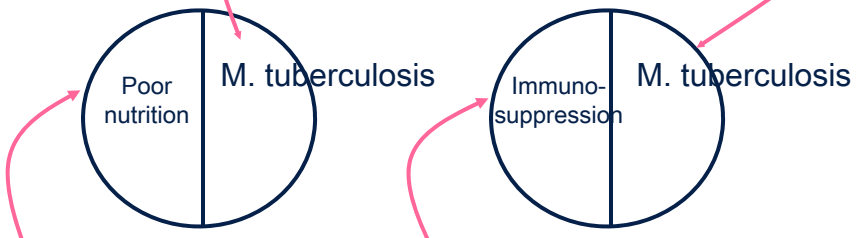
Sufficient Cause 1 Sufficient Cause 2

63

UCSF

Example: Tuberculosis

Necessary but not sufficient



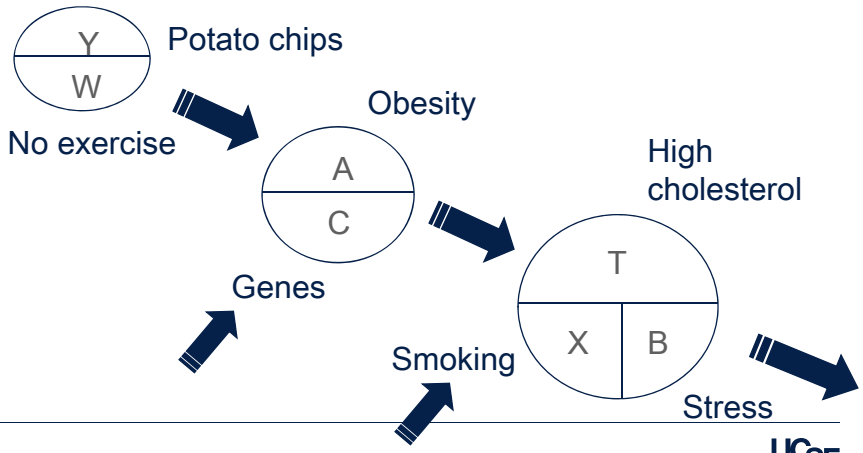
Sufficient Cause 1 Sufficient Cause 2

Neither necessary nor sufficient

64

UCSF

"Causing" a myocardial infarction



65

UCSF

Advantages of the Rothman's model

- Analogies could be drawn with a multistage carcinogenesis model, where initiators, promoters and catalysts could be conceptualized as component causes

66

UCSF

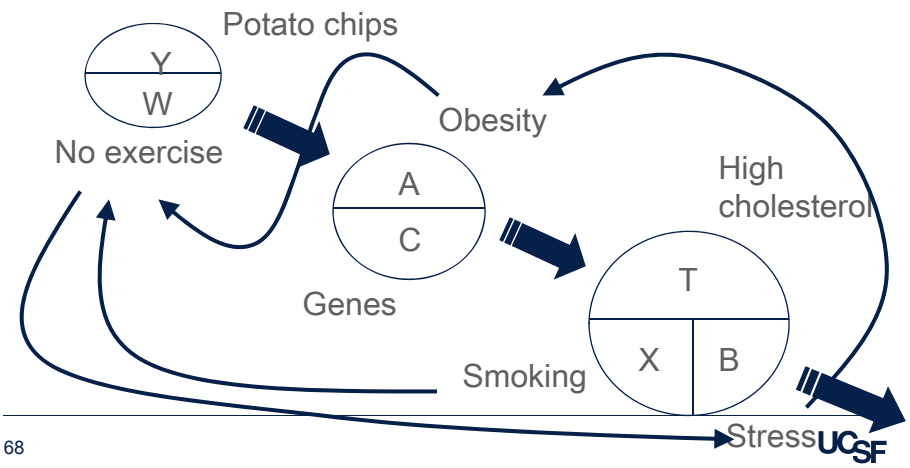
Limitations of Rothman's model

- Omits discussion of origins of causes, focuses on proximal causes and ignores different induction periods for different component causes
- Specific components but not linkages among them
  - Ignores indirect effects (effects of some component causes mediated by other component causes in the model)
- Causes of disease in individuals but not in populations
- Does not consider factors that control distribution of risk factors
- Ignores dynamic non-linear relations

67

UCSF

"Causing" a myocardial infarction



68

UCSF

UCSF

Counterfactual model (potential-outcome model)

a causal effect is a counterfactual contrast between the outcomes of a single unit under different treatment possibilities

Sander Greenland **Hal Morgenstern**

Neyman 1923
Rubin 1974
Greenland and Morgenstern *Annu Rev Public Health* 2001

69

UCSF

Counterfactual model

- Ideal comparison to obtain a measure of effect would be of study subjects with themselves in both an exposed and an unexposed state
- One of the two conditions in the definitions of the effect measures must be contrary to fact – exposure or treatment vs. a reference condition
- Specifies what would happen to individuals or populations under alternative possible patterns of interventions or exposures

70

UCSF

Advantages of the counterfactual model

- Can be applied at mechanistic, individual or population levels of analysis
- Forces researchers to think about operational definitions of cases and controls, sampling schemes, and other important design questions
- Provides direct representations of epidemiologic concepts of confounding, bias, interaction

Limitations of the counterfactual model

- Effects are defined only for comparisons of treatment levels
- Causes refer to factors that can be potentially manipulated, such as drug treatments, but not to fixed personal attributes such as gender and race
- Implicit in most discussions of potential outcomes is that the outcome for a given unit under a specific treatment does not depend on the treatment given to any other unit, i.e., *the independent outcomes assumption*. This assumption is likely to be violated when the outcome is contagious or the exposure represents a set of social conditions

Limitations of the counterfactual model

- In considering causes of past events, we invoke consideration of distributions for events that never occurred and hence cannot be observed
- Need to make assumptions that are untestable given ethical considerations or limitations of current knowledge and technology:
  - exchangeability (no confounding)
  - stability (i.e., consistency, irrespective of the number of ways in which treatment X could have been assigned, it would have resulted in the same observed outcome)
  - positivity (a non-zero probability of receiving every level of exposure for every combination of values of exposure and covariates that occur among individuals in the population).
  - correct model specification
  - independence of unit-specific susceptibilities or responses (non-interference)
  - various distributional assumptions ➢ **W#6 Confounding lecture**

73 Cole & Frangakis *Epidemiology* 2009

UCSF

Utility of the counterfactual model

- Counterfactual approaches to causal inference emphasize the importance of randomization in assuring identifiability of causal effects.
- In observational studies, no such assurance is available, and issues of confounding become paramount.
- Assumption of exchangeability of response distributions under homogeneous treatment assignment, which is met when treatment is successfully randomized or, more generally, when treatment assignment is independent of the potential outcomes

74

UCSF

Utility of the counterfactual model

Effect measures vs. measures of association

- Effect is:
  - The endpoint of the causal mechanism
  - Change in a population characteristic that is caused by the factor being at one level versus another
- Effect measures:
  - Can never achieve counterfactual ideal
  - Logically impossible to observe the population under both conditions
- Measures of association
  - Compares what happens in two distinct populations
  - Constructed to equal the effect measure of interest
  - Absolute: differences in occurrence measures (rate or risk difference)
  - Relative: ratios of occurrence measures (rate or risk ratio, relative risk, odds ratio)

Comparison of absolute and relative effect measures

Measure	Numerical Range	Dimensionality
Risk difference	**[-1, +1]**	**None**
Risk ratio	**[0, ∝]**	**None**
Incidence rate difference	**[- ∝, + ∝]**	**1/Time**
Incidence rate ratio	**[0, ∝]**	**None**

Interrelationships between relative measures of disease occurrence

- Rare disease assumption (Cornfield 1951)
  - If disease is rare, the odds ratio approximates the risk ratio

OR=odds of disease among exposed/odds of disease among unexposed

=(A/B) / (C/D)

	Disease	
	+	-
Exposure +	A	B
Exposure -	C	D

RR=risk in exposed/ risk in unexposed

=(A/(A+B)) / (C/(C+D))

➢**We will come back to examine this "collapsibility" assumption in Week #8 Confounding**

77

UCSF

Interrelationships between relative measures of disease occurrence

Exposure only negligibly affects the person-time at risk (T1≈T0)
IRR=incidence rate among exposed/ incidence rate among unexposed

=(IR1xT1) / (IR0xT0)

RR=risk in exposed/ risk in unexposed

=(A/(A+B)) / (C/(C+D))=R1/R0

78

UCSF

Interrelationships between relative measures of disease occurrence

- If R1>R0, then A/C>1, i.e. OR overestimates association and is larger than RR and further away from the null
- If R1>R0, then T1<T0 and T1/T0<1, i.e. IR1/IR0>R1/R0,

i.e. rate ratio falls between the risk ratio and the odds ratio

$$1 < RR < IRR < OR$$

	Disease	
	+	-
Exposure +	**A**	**B**
Exposure -	**C**	**D**

RGL2008, Ch 4

UCSF

Measures of attributable risk

- **Formula for attributable fraction (excess fraction):**

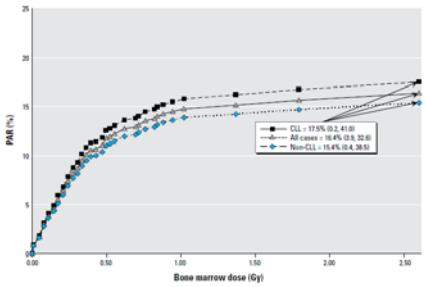
For dichotomous exposure: AF=(R1 – R0)/ R1=(RR-1)/RR

For categorical (n>2) exposure: Σi (AFi x Pi)

- NB! Different from **etiologic fraction**, which refers to Rothman's causal heuristic and adds up to more than 100%

- **Formula for population attributable risk:**

PAR = Σk Pk × (RRk – 1)/Σk Pk × RRk, where k = 0, 1, … , 100, and Pk and RRk are the proportion and model-based estimates of RR at the kth percentile dose level.



Zablotska et al. *EHP* 2013

UCSF

UCSF

Graphical causal models

- Provide a unified framework for evaluating design and analysis strategies for any causal question under any set of causal assumptions
- intuitive device for deducing the statistical associations implied by causal relations. Conversely, given a set of observed statistical relations, a researcher armed with causal graph theory can systematically characterize all causal structures compatible with the observations.

Example:
Comparison of mortality in Sweden and Panama

- Our prediction:
  - Standard of living in Sweden is generally higher than in Panama
  - Panama has more limited health care and higher poverty rates compared to Sweden
  - Proportion of the population that dies each year is higher in …?

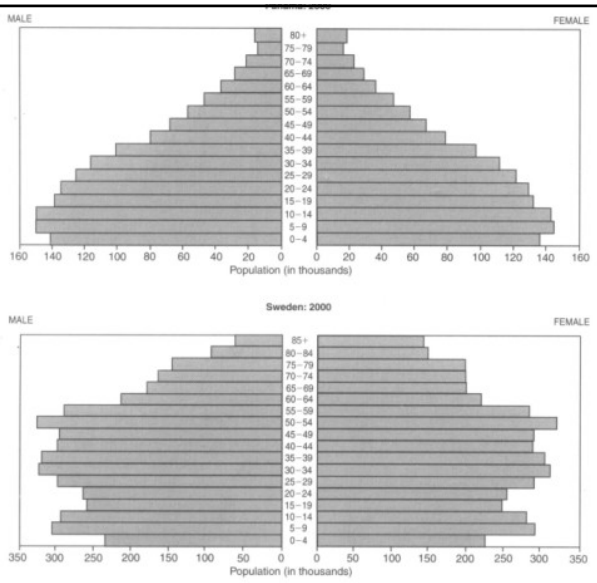


Figure 1–1. Age distribution of the populations of Panama and Sweden (population

UCSF

Example:

Comparison of mortality in Sweden and Panama

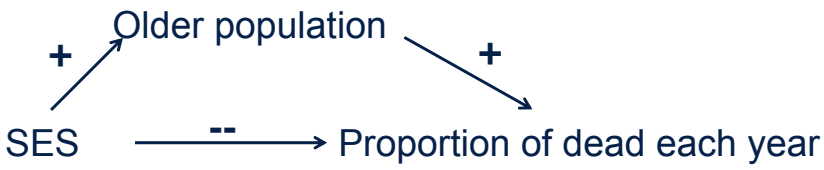
- Our prediction:
  - Standard of living in Sweden is generally higher than in Panama
  - Panama has more limited health care and higher poverty rates compared to Sweden
  - Proportion of the population that dies each year is higher in …?
- Actual findings:
  - Death rates are lower for people of the same age in Sweden
  - In both countries older people die at a greater rate than younger people
  - Proportion of older people is greater in Sweden

UCSF

UCSF

Causal diagrams

- Provide a unified framework for evaluating design and analysis strategies for any causal question under any set of causal assumptions



85

Directed acyclic graphs (DAGs)

- Much more about DAGs in Week #2

86

Complex systems methods (e.g., agent-based modeling)

- Complex systems methods: system dynamics, network analysis, agent-based modeling

  - For many diseases, there are numerous factors at different levels of influence (biological, behavioral, group and macro-social)
  - These factors frequently influence one another and, in addition, are sometimes influenced by the outcomes

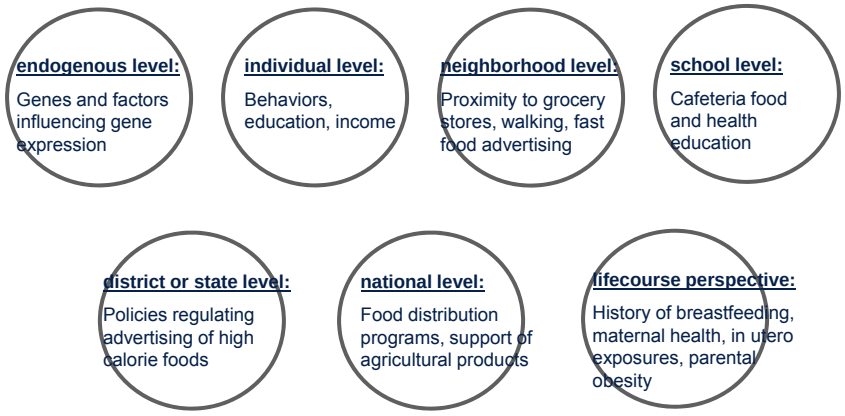
↓

Multidimensional approach

87 Galea et al. *Int J Epidemiol 2010*

UCSF

Example: obesity epidemic in the U.S.



88 Galea et al. *Int J Epidemiol 2010*

UCSF

Determining causality using different causal models

- Sufficient component causes model – Are higher-level factors true causes?
- Counterfactual model – Will changes in some factors always lead to changes in the outcome?
- Multi-level (hierarchical) regression modeling – Could they take account of dynamic and reciprocal relationships between factors and outcomes?
- Complex systems dynamic approach – Allows feedback loops, interrelations among factors, and discontinuous non-linear relations

89

UCSF

Agent-based modeling (ABM)

1. Specify the rules for individual agents (usually individuals) and how they interact with each other
2. Run multiple simulations to observe how these individual-level rules translate into patterns at the population level
3. Compare outcomes obtained from numerous hypothetical scenarios

Agents:

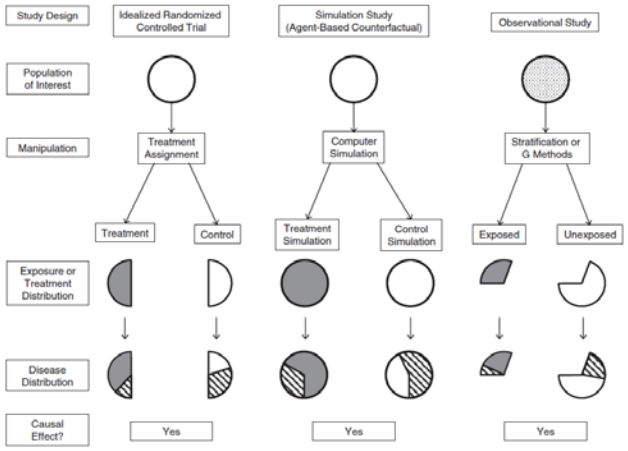
- could be adaptive, so that their behavior changes over time
- could interact (based on proximity or on a social network structure)
- could be clustered
- feedback loops between agents
- could have interference (one person's exposure affects the outcome of others)
- could be affected by threshold dynamics

90

Galea et al. *Int J Epidemiol 2010*
Marshall & Galea *Am J Epidemiol* 2014

UCSF

Comparison of causal model in estimation of average causal effects



91 Marshall & Galea *Am J Epidemiol* 2014

UCSF

ABM example

- Agent's diet depends on availability of good food stores, education, parental and friend diet, and genetics
- Agent's diet influences BMI, with a feedback loop
- Intervention: Increase in the level of investment in attracting good food stores in 100 neighborhoods
- Simulation: Strong network ties = agent's diet depends 90% on the diets of their friends
- Modeled using Recursive Porous Agent Simulation Toolkit (REPAST) software

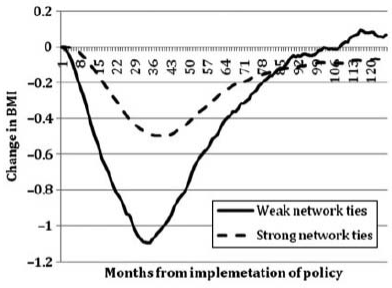


Figure 1 Agent-based-modelling simulation of population changes in BMI subsequent to the implementation of a policy to attract better food stores to local neighbourhoods, stratified by populations characterized by strong and weak network ties

92 Galea et al. *Int J Epidemiol 2010*

UCSF

Advantages of the ABM

- Eliminates "compartmental" causation models attached to just one level
- Could be used to examine the effects of multiple exposures that interact in non-linear and dynamic ways to affect population-level health outcomes – could solve many contemporary public health problems (the end of "fiber yarn"???)
- Overcomes limitations of the previous models, e.g., allows non-linearity in relationships, adaptivity, feedback loops, contextual effects, social processes (e.g., the diffusion of health information), multidirectionality, interference
- Multiple ABM outcomes could be interpreted as counterfactual outcomes

93

UCSF

Limitations of the ABM

- Theory and tools are not well-developed
- Strongly depends on assumptions
- Validity of causal conclusions depends on the extent to which the models represent the true world
- Where does the data come from?
- We already have a g-estimation method to do the same using a single dataset
  - **More in Week #8 Confounding**

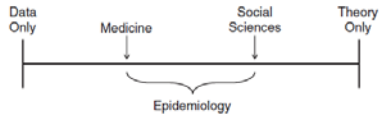


Figure 1. The relative position of several scientific disciplines along the causal inference spectrum according to the relative weights of data and theory.

Marshall & Galea *Am J Epidemiol* 2014
Diez Roux *Am J Epidemiol* 2014
Hernan *Am J Epidemiol* 2014

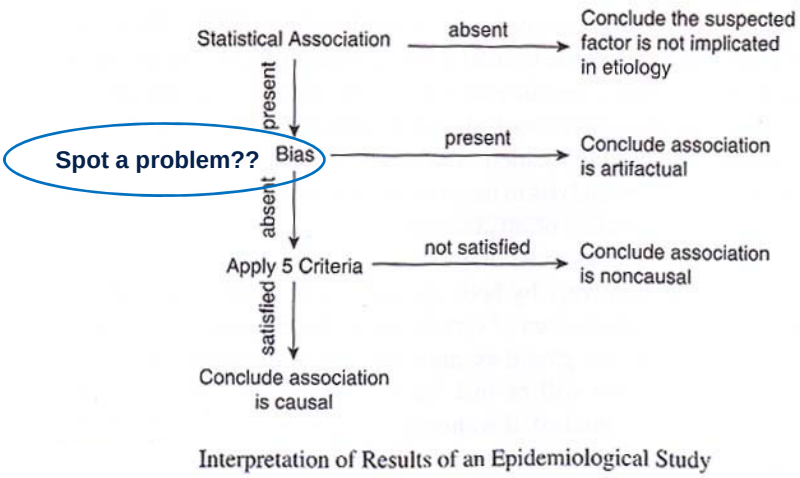
94

UCSF

Practice of causal inference

1. Is the observed association valid? Is it true?
   - Association appears causal but is due to:
     - Bias or systematic error (misclassification of E or D)
     - Confounding (other variable causes the D and this variable correlates with E)
     - Chance or random error (just this once)
2. Did the exposure actually cause the disease?
   - Use causal guidelines to decide if association is truly causal
   - The most important is temporality

Summary: When is an association causal?



Interpretation of Results of an Epidemiological Study

Ethics and the public health balance

- When is there enough evidence to say something is a “cause”?
- When should we decide that something is a cause and act on it?
- Does “first do no harm” always apply at the population level?
- Are there different guidelines for solutions where we have to DO something vs. solutions where we try to remove something?

Summary

- Causal inference is the central goal of many, if not most, empirical investigations in the fields of medicine, epidemiology and public health
- Causal inference is neither quick nor easy
- No single study is sufficient in establishing causal inference
- Causal inference requires critical judgment and interpretation
- We could rarely “prove” causal associations
- Existing models (and heuristics) are the necessary simplifications - simple, efficient rules which people often use to form judgments about causal associations. Causal models usually focus on one aspect of a complex problem and ignore others, which could lead to errors
- Careful design of empirical investigations using inductive reasoning could prevent many errors and biases

"And it was so typically brilliant of you to have invited an epidemiologist.

99 November 26, 2001, *The New Yorker.*

UCSF

UCSF
University of California
San Francisco