I would rather discover one causal law than be King of Persia.
— Democritus

We have knowledge of a thing only when we have grasped its cause.
— Aristotle, Posterior Analytics

We do not have knowledge of a thing until we have grasped its why, that is to say, its cause.
— Aristotle, Physics
Questions on Causation

Relevant questions about causation:
- the philosophical meaningfulness of the notion of causation
- deducing the causes of a given effect
- understanding the details of causal mechanism

In this class we focus on measuring the effects of causes – a place where statistics, which is concerned with measurement, has most contributions to make.
Association vs. Causation

- The research questions that motivate most studies in statistics-based sciences are causal in nature.

- Standard statistical analysis is to infer **associations** among variables, based on which may do some **prediction**

- Causal analysis is one step further: it is about **counterfactual prediction**, predict what would have happened to the same units/subjecs had they were exposed to a different (counterfactual) condition

- In most cases, **association does not imply causation**
Causal Inference

- *How to make the leap from association to causation?*

- **Key:** *causal assumptions* - structural and/or modeling

- Causal inference is about
  1. build a framework and define causal effects under general scenarios
  2. specify assumptions under which one can declare/identify causation from association
  3. assess the sensitivity to the causal assumptions and find ways to mitigate
Notations

- Treatment (e.g. intervention, exposure) $Z$: for illustration we will mostly focus on binary treatments
- Outcome (e.g. disease status) $Y$
- Observed covariates or confounders $X$
- Unobserved covariates or confounders $U$

Examples of question of interest
- Causal effect of exposure on disease
- Comparative effectiveness research: whether one drug or medical procedure is better than the other
- Program evaluation in economics and policy
Three Type of Biases
(Hernan and Robins, 2020)

▶ Types of biases
  ▶ Confounding bias
  ▶ Selection bias
  ▶ Measurement bias

▶ All three have strong connections to missing data

▶ Will mostly address confounding bias and (if time allows) selection bias in this course
  ▶ germane to observational studies, but also in randomized trials
Confounding

- Confounding (or common cause) is the main complication/hurdle between association and causation

- Two Directed Acyclic Graphs

  Causal relationship: \[ Z \rightarrow Y \]

  Confounding: \[ Z \leftarrow \text{confounder} \rightarrow Y \]
Examples of Confounding

- Ice cream consumption and number of people drown. Confounder: temperature
- Medical treatment and patient outcome. Confounders: age, sex, other complications
- Education and income. Confounder: SES of family
- An extreme example of confounding is Simpson’s paradox: confounder reverses the sign of the correlation between treatment and outcome
Simpson’s paradox: Kidney Stone Treatment
(Charig et al., BMJ, 1986)

- Compare the success rates of two treatments for kidney stones
- Treatment A: open surgery; treatment B: small puncture

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small stones</td>
<td>93% (81/87)</td>
<td>87% (234/270)</td>
</tr>
<tr>
<td>Large stones</td>
<td>73% (192/263)</td>
<td>69% (55/80)</td>
</tr>
<tr>
<td>Both</td>
<td>78% (273/350)</td>
<td>83% (289/350)</td>
</tr>
</tbody>
</table>

- What is the confounder here? Severity of the case
Simpson’s paradox or Yule-Simpson effect
(K Pearson et al., 1899; Yule, 1903; Simpson, 1951)

- Simpson’s paradox: a trend appears in different groups of data but disappears or reverses when these groups are combined

Mathematically, it is about conditioning
- A special case of the ecological fallacy
- It may or may not be due to confounding
- Another well-known example is the Berkeley admission gender bias (Bickel et al., Science, 1976)
A Classic Example—Smoking and Lung Cancer
Doll and Hill (1950 BMJ)

- Smoking-cancer association
- Case-control study of lung cancer
- Risk ratio \( \approx \) odds ratio, is roughly 9 even after adjusting for observed covariates:
  \[
  \text{RR}_{ZY}^{\text{obs}} = \frac{\Pr(Y = 1 \mid Z = 1)}{\Pr(Y = 1 \mid Z = 0)} \approx 9
  \]

- Does smoking cause lung cancer?
- Box (2013) stopped smoking after seeing Doll and Hill (1950)

Figure: Sir Austin Bradford Hill (1897–1991)
A Classic Example—Smoking and Lung Cancer

Figure: Sir Ronald Aylmer Fisher (1890–1962)

- Association does not imply causation
- “Common cause” (Reichenbach 1956, Fisher 1957 BMJ)
- Fisher (1957 BMJ): *cigarette-smoking and lung cancer, though not mutually causative, are both influenced by a common cause, in this case the individual genotype.*
Common cause hypothesis

- Assume Fisher is right
- The smoking-gene association must satisfy:
  \[ RR_{ZU} \geq RR_{ZY} \approx 9 \]
- Such a genetic confounder is too strong to be realistic.
- Association must be due to causal.
- We will revisit this example later.
Frameworks for Causal Inference

- The purpose is to construct a model or a framework that is complex enough to allow us to formalize basic intuitions concerning cause and effect.

- Two commonly used frameworks
  - The *potential outcome framework*, also known as the counterfactual framework, or the Neyman-Rubin Causal Model (Neyman, 1923; Rubin, 1974; Imbens and Rubin, 2015; Hernan and Robins, 2020)
  - The *causal diagram* framework (Pearl, 2009)
  - Mathematically the two frameworks are connected (Richardson and Robins, 2013), but each has different established goals, tools and applicable areas

- This class focuses on the potential outcome framework, and will occasionally draw directed acyclic graphs (DAGs) for simple illustration.
The Potential Outcome Framework is arguably the most widely used causal framework across many disciplines, e.g., medicine, health care, policy, social sciences.

Under the PO framework, causal inference is viewed as a problem of missing data with explicit mathematical modeling of the assignment mechanism as a process for revealing the observed data.

Rooted in the statistical work on randomized experiments by Fisher (1918, 1925) and Neyman (1923), as extended by Rubin (1974, 1978) and subsequently by many others to apply to nonrandomized studies and other forms of inference.
Potential Outcome Framework

- No causation without manipulation - “cause” must be (hypothetically) manipulatable, e.g., intervention, action, treatment

- Gender, time and age are not well defined “causes” under the PO framework

- Three integral components (Rubin, 1978):
  - potential outcomes corresponding to the various levels of a treatment or manipulation (no causation without manipulation)
  - assignment mechanisms
  - a (Bayesian) model for the science (the potential outcomes and covariates)
Potential Outcome Framework: Basic Concepts

- **Unit**: The person, place, or thing upon which a treatment will operate, at a particular time (Note: A single person, place, or thing at two different times comprises two different units)

- **Treatment**: An intervention, the effects of which (on some particular measurement of the units) the investigator wishes to assess relative to no intervention (i.e., the control)

- **Potential Outcomes**: The values of a unit’s measurement of interest after (a) application of the treatment and (b) non-application of the treatment (i.e., under control)

- **Causal Effect**: For each unit, the comparison of the potential outcome under treatment and the potential outcome under control
For a single unit, let \( Y(0) \) denote the outcome given the control treatment, and \( Y(1) \) the outcome given the active treatment.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Initial Headache</th>
<th>Potential Outcomes</th>
<th>Causal Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>X 80</td>
<td>( Y(\text{asp}) ) -50</td>
<td>( Y(\text{asp}) - Y(\text{not}) )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>-50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td></td>
</tr>
</tbody>
</table>

**Gain Scores**

<table>
<thead>
<tr>
<th>Unit</th>
<th>Initial Headache</th>
<th>Potential Outcomes</th>
<th>Causal Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>you</td>
<td>X 80</td>
<td>( Y(\text{asp}) - X ) -50</td>
<td>( [Y(\text{asp}) - X] - [Y(\text{not}) - X] )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-55</td>
<td>-50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-5</td>
<td></td>
</tr>
</tbody>
</table>
The fundamental problem of causal inference: We can observe at most one of the potential outcomes for each unit, the other(s) are missing/counterfactual.

Causal inference under the potential outcome framework is essentially a missing data problem.

To identify causal effects from observed data, under any mathematical framework, one must make assumptions (structural or/and stochastic).
Multiple Units

- We must rely on multiple units exposed to different treatments to make causal inferences
  - observe the same physical object subject to different treatment levels at different points in time
  - observe different physical units at the same time
- By itself, however, the presence of multiple units does not solve the problem of causal inference

<table>
<thead>
<tr>
<th>You take:</th>
<th>Asp</th>
<th>Not</th>
<th>Asp</th>
<th>Not</th>
</tr>
</thead>
<tbody>
<tr>
<td>I take:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 1=you</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1=you</td>
<td>$Y_1([A, A]) = 0$</td>
<td>$Y_1([N, N]) = 100$</td>
<td>$Y_1([A, N]) = 50$</td>
<td>$Y_1([N, A]) = 75$</td>
</tr>
<tr>
<td>2=me</td>
<td>$Y_2([A, A]) = 0$</td>
<td>$Y_2([N, N]) = 100$</td>
<td>$Y_2([A, N]) = 100$</td>
<td>$Y_2([N, A]) = 0$</td>
</tr>
</tbody>
</table>
SUTVA

Assumption 1: Stable Unit Treatment Value Assumption (SUTVA) (Rubin, 1980):

- SUTVA includes two assumptions: (1) no interference, (2) no different versions of a treatment (also known as consistency, Cole and Frangakis, 2009)

- Seems trivial, actually very strong assumptions

- Under SUTVA, we have

  - If $Z_i = 1$ then $Y_i = Y_i(1)$

  - If $Z_i = 0$ then $Y_i = Y_i(0)$

- Equivalently: $Y_i = Z_iY_i(1) + (1 - Z_i)Y_i(0)$

- SUTVA connects the intervention we see ($Z$), with the causal intervention of interest ($w$):
Interference

- **Interference**: the potential outcome $Y_i(z)$ for an individual $i$ depends on what treatment other people receive.

- Examples: vaccination, advertising, infectious diseases, social networks, agricultural experiments.

- There are lots of possible $Y_i(z)$, *depending on what happens to other people*.

- When in the presence of interference, other assumptions required for causal inference (e.g., Rosenbaum 2007; Hudgens and Hollaran 2008) – a hot topic lately.
Multiple Versions of Treatment

- **Multiple Versions of Treatment**: Sometimes action $Z$ doesn’t have a clear meaning, as it has many versions.

- **Examples**
  - Dose
  - Surgeon
  - Treatment history
  - Standard of care (as a reference)

- There are lots of possible $Y_i(Z)$, *depending on what version gets selected*. 
Additional Implicit Assumptions

- Well-defined *outcome* of interest (consistency)
  - Clinical definition
  - Ascertainment methods
  - Duration of follow-up

- *Target Population*: a well-defined *population* of individuals whose outcomes are going to be compared
Basic Setup

- Data: a random sample of $N$ units from a target population
- A treatment with two levels: $w = 0, 1$
- For each unit $i$, we observe the (binary) treatment status $Z_i (= 0, 1)$, a vector of $p$ covariates $X_i = (X_{i1}, \ldots X_{ip})$, and an outcome $Y_{i}^{obs}$ (or simply denoted as $Y_i$ later)
- For each unit $i$, two potential outcomes $(Y_i(0), Y_i(1))$: the outcomes under the two values of the treatment, at most one of which is observed
- Potential outcomes and assignments jointly determine the values of the observed and missing outcomes:

$$Y_{i}^{obs} \equiv Y_i(Z_i) = Z_i \cdot Y_i(1) + (1 - Z_i) \cdot Y_i(0)$$

$$Y_{i}^{mis} \equiv Y_i(1 - Z_i) = (1 - Z_i) \cdot Y_i(1) + Z_i \cdot Y_i(0)$$
Causal Estimands: Difference

- Conditional average treatment effect (CATE), also known as individual treatment effect (ITE): conditional on a covariate value

  \[ \tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) | X = x] \]

- Average treatment effect (ATE):

  \[ \tau = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}_x[\tau(x)]. \]

- Average treatment effect for the treated (ATT):

  \[ \tau = \mathbb{E}[Y_i(1) - Y_i(0) | Z_i = 1]. \]
Causal Estimands: Ratio

- Individual treatment effect in ratio: $r_i = Y_i(1)/Y_i(0)$
- Population treatment effect in ratio: $r = \mathbb{E}[Y_i(1)]/\mathbb{E}[Y_i(0)]$
- Relative causal effect (lift): $l = \mathbb{E}[Y_i(1) - Y_i(0)]/\mathbb{E}[Y_i(0)]$
- Continuous outcomes are common in theoretical literature and applications social sciences
- Binary outcomes are more common in medical and health studies

Note: ratio is not an average of individual causal effect unless the effect is homogeneous
Causal Estimands: Binary outcomes

- For binary outcomes, we have specific estimands and terminology
- Let $\mu_1 = \Pr(Y(1) = 1)$ and $\mu_0 = \Pr(Y(0) = 1)$, often known as causal risks in medical research.
- Three estimands:
  - Causal risk difference (equivalent to ATE): $\tau_{RD} = \mu_1 - \mu_0$
  - Causal risk ratio: $\tau_{RR} = \frac{\mu_1}{\mu_0}$
  - Causal odds ratio: $\tau_{OR} = \frac{\mu_1/(1-\mu_1)}{\mu_0/(1-\mu_0)}$
- These estimands are different from their observed counterpart (denote $m_1 = \Pr(Y^{obs} = 1|Z = 1)$ ($m_0 = \Pr(Y^{obs} = 1|Z = 0)$)
  - Observed risk difference: $m_1 - m_0$
  - Observed risk ratio: $\frac{m_1}{m_0}$
  - Observed odds ratio: $\frac{m_1/(1-m_1)}{m_0/(1-m_0)}$
- We will explain why they are different later
Potential Outcomes

- Causal effects defined by potential outcomes, not model parameters

- Why bother to introduce the hypothetical potential outcomes?
  - separation of the intervention and quantities of interest
  - randomness comes solely from the treatment
  - analysis of experiments is driven by design

- Obviously these estimands are not identifiable without further assumptions

- So, what assumptions do we need?
### Perfect Doctor

<table>
<thead>
<tr>
<th>Potential Outcomes</th>
<th>Observed Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y(0)$</td>
</tr>
<tr>
<td>$Y(0)$</td>
<td>13</td>
</tr>
<tr>
<td>$Y(1)$</td>
<td>6</td>
</tr>
<tr>
<td>$Y(0)$</td>
<td>4</td>
</tr>
<tr>
<td>$Y(1)$</td>
<td>5</td>
</tr>
<tr>
<td>$Y(0)$</td>
<td>6</td>
</tr>
<tr>
<td>$Y(1)$</td>
<td>6</td>
</tr>
<tr>
<td>$Y(0)$</td>
<td>8</td>
</tr>
<tr>
<td>$Y(1)$</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True averages</th>
<th>7</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed averages</td>
<td>5.4</td>
<td>11</td>
</tr>
</tbody>
</table>
Perfect Doctor: Comments

- The simple difference-in-means estimator does not return a valid estimate of the true causal effect. Why?

- Key: the assignment mechanism (Rubin, 1978) – the probabilistic rule that decides which unit gets assigned to which treatment

\[ p(Z_i = 1|X_i, Y_i(0), Y_i(1)) \]

- Is the assignment mechanism in the perfect doctor example random? (in fact, what does random assignment mean?)

- The assignment depends on both \( Y_i(0) \) and \( Y_i(1) \) for each unit (of course, in reality we will unlikely to have any such perfect doctor)

- The key identifying assumptions in causal inference are on the assignment mechanism
Unconfounded Assignment

- *Unconfounded Assignment*: an assignment is unconfounded if the assignment mechanism does not depend on the potential outcomes (and only on the pre-treatment covariates):

\[ p(Z_i = 1|X_i, Y_i(0), Y_i(1)) = p(Z_i = 1|X_i) \]

- \(p(Z_i = 1|X_i)\) is also known as the propensity score (Rosenbaum and Rubin, 1983)

- Is the assignment mechanism unconfounded
  - in randomized experiments?
  - in the perfect doctor example?
  - in the smoking and lung cancer study?
Unconfounded Assignment

- **Unconfounded Assignment**: an assignment is unconfounded if the assignment mechanism does not depend on the potential outcomes (and only on the pre-treatment covariates):

\[ p(Z_i = 1|X_i, Y_i(0), Y_i(1)) = p(Z_i = 1|X_i) \]

- \( p(Z_i = 1|X_i) \) is also known as the propensity score (Rosenbaum and Rubin, 1983)

- Is the assignment mechanism unconfounded
  - in randomized experiments? Yes, and it is known and controlled by investigators
  - in the perfect doctor example? No, because it depends on both \( Y_i(1) \) and \( Y_i(0) \)
  - in the smoking and lung cancer study? We do not know, as in most observational studies, the assignment mechanism is usually unknown and uncontrolled. To make causal inference, we have to make (often strong) assumptions about the assignment mechanism
Ignorable Assignment

- An assignment mechanism is ignorable if it does not depend on the missing outcomes:
  \[ p(Z_i = 1|X_i, Y_i(0), Y_i(1)) = p(Z_i = 1|X_i, Y_{i\text{obs}}) \]

- The terminology is corresponding to the missing data mechanism in the missing data literature (Rubin, 1976).

- The meaning of “ignorability" becomes more apparent when we talk about Bayesian inference of causal effect – meaning the assignment mechanism drops out from the data likelihood in estimating the causal effects.

- An unconfounded assignment is always ignorable, but not vice versa (one example is the sequential randomized experiment).

- In most cases in practice, the difference between unconfoundedness and ignorability is negligible, and these two are used exchangeably.
Outline of the Class

The class material is (mostly) organized by the classification of assignment mechanisms

- **Randomized Experiments**: the assignment mechanism is known, controlled, and random (stronger than unconfounded)
- **Observational studies**: the assignment mechanism is unknown and uncontrolled, but often assumed to be unconfounded conditional on observed covariates or unobserved quantities
  - Cross-sectional data: treatment at one time point
  - Longitudinal data: treatment at multiple time points (sequentially ignorable)
  - Panel data: treatment at one time point, comparative case studies (often different assumptions than unconfoundedness)
- **Natural/quasi-experiments**: usually involve some type of unconfoundedness assumptions (can based on unobserved quantities)
  - instrumental variables
  - principal stratification (latent ignorable)
  - regression discontinuity (local randomized)
Methods and Modes of Inference

- Two overarching methods
  - Imputation: impute the missing potential outcomes (model-based or matching-based)
  - Weighting: weight (often function of the propensity scores) the observed data to represent a target population

- Three modes of inference
  - Frequentist: imputation, weighting, motivated by consistency, asymptotic normality, (semiparametric) efficiency, etc.
  - Bayesian: modeling and imputing missing potential outcomes based on their posterior distributions
  - Fisherian randomization: combine randomization tests with Bayesian methods, unique to randomized experiments
Causal Inference vs. Missing Data

- Under PO framework, causal inference is a missing data problem.

- A broad parallel between the classification of assignment mechanisms in causal inference and the classification of missing data mechanisms (Ding and Li, 2018, Stat Sci)
  - completely randomized experiments $\iff$ missing completely at random (MCAR)
  - observational studies with unmeasured confounding $\iff$ missing not at random (MNAR)
  - ignorable assignment mechanisms MAR $\iff$ missing at random (MAR)

- But ignorable assignment mechanism can be generalized to a richer class of assignment mechanisms (sequential ignorable, locally ignorable, etc).
Selection Bias

- The term “selection bias" encompasses various biases that arise from the procedure by which individuals are selected into the analysis (Hernan and Robins, 2020)

- Examples
  - Differential loss to follow-up (informative censoring)
  - Missing data bias
  - Healthy worker bias
  - Self-selection bias
Selection Bias

- Selection bias (or common effect) frequently arises in estimating causal effect with time-to-event data (survival analysis).

- Define \( C \) as the censoring indicator (\( C = 1 \) if loss to follow up or death, and \( C = 0 \) otherwise).

- Two Directed Acyclic Graphs (DAGs)
  
  Causal relationship: 

  Death (common effect): 

  \[
  \begin{align*}
  Z & \rightarrow Y \\
  Z & \rightarrow \text{death } C \\
  \text{death } C & \rightarrow Y
  \end{align*}
  \]
A Hypothetical Example
(Hernan and Robins, 2020)

- Consider a randomized study to estimate the effect of folic acid supplements \( Z \) given to pregnant women shortly after conception on the fetus’s risk of developing a cardiac malformation \( Y \) (1: yes, 0: no) during the first two months of pregnancy.
- Variable \( C \) represents death before birth.
- A cardiac malformation increases mortality \((Y \rightarrow C)\), and folic acid supplementation decreases mortality by reducing the risk of malformations other than cardiac ones \((arrow \ from \ Z \rightarrow C)\).
- The study was restricted to fetuses who survived until birth.
  - conditioning on the common effect \((or \ C = 0)\).
  - a harmful effect may be found when the truth is null \((more \ cardiac \ malformation \ remain \ in \ Z = 1 \ group)\).
Conditionally Noninformative Censoring

Mathematically, the problem is stated as

\[
\frac{P(Y_i = 1 \mid Z_i = 1, C_i = 0)}{P(Y_i = 1 \mid Z_i = 0, C_i = 0)} = \frac{P(Y_i(1) = 1 \mid C_i = 0)}{P(Y_i(0) = 1 \mid C_i = 0)} \neq \frac{P(Y_i(1) = 1)}{P(Y_i(0) = 1)}.
\]

However, similar to unconfoundedness, if we are able to measure a set of risk factors that explain the censoring mechanism, one could recover the marginal causal risk ratio

\textit{Conditionally Noninformative Censoring}: the censoring is conditionally uninformative if it does not depend on the potential outcomes:

\[
P(C_i = 1 \mid X_i, Y_i(0), Y_i(1), Z_i) = P(C_i = 1 \mid X_i, Z_i)
\]

Again, a form of conditional exchangeability assumption
Selection Bias and Generalizability

- Sometimes the term “selection bias" is used to refer to lack of generalizability of measures of effect.

- Here the definition of target population becomes more explicit:
  - Selection of patients into randomized studies
  - Sample average treatment effect (SATE)
  - Population average treatment effect (PATE)

- Requires a conditional exchangeability assumption on sampling or study participation (Cole and Stuart, 2010)
Measurement Bias

- In many studies, we made the implicit assumption that all variables were perfectly measured, which may be unrealistic
  - imperfectly measured exposure or treatment
  - imperfectly measured outcomes
  - imperfectly measured confounders

- We say that there is measurement bias when the association between treatment and outcome is weakened or strengthened as a result of the process by which the study data are measured

- Measurement error correction often requires additional data information to relate mismeasured variables with gold standard (exchangeability assumptions will become useful to recover the true population relationship)

- Not the focus of this class
A Few Comments

- Causal effects, in their definition, do not relate to probability distribution for subjects “who got different treatments”, or to coefficients of models.

- We do not assume necessarily that the P.O.s are fixed. In fact we could consider an inherent status as random, but that is not important for the scale we discuss here.

- We allow that the potential outcomes depend on other factors that relate to that person at the time and place when the treatment is given, and through the follow-up time where the P.O.s would be observed.

- Not a before-after comparison.
FL: In the RCM, cause/intervention should always be defined before you start the analysis. In other words, the RCM is a framework to investigate the "effects of a cause," but not the "causes of an effect." Some criticize this as a major limitation. Do you regard this as a limitation? Do you think it is ever possible to draw inference on the causes of effects from data, or is it, per se, an interesting question worth further investigation?

Rubin: I regard "the cause" of an event topic as more of a cocktail conversation topic than a scientific inquiry, because it leads to an essentially infinite regress. Someone says, "He died of lung cancer because he smoked three packs a day"; then someone else counters, "Oh no, he died of lung cancer because both of his parents smoked three packs a day and, therefore, there was no hope of his doing anything other than smoking three packs a day"; then another one says, "No, no, his parents smoked because his grandparents smoked - they lived in North Carolina where, back then, everyone smoked three packs a day, so the cause is where the grandparents lived," and so on. How far back should you go? You can't talk sensibly about the cause of an event; you can talk about "but for that cause (and there can be many 'but for's), what would have happened?" All these questions can be addressed hypothetically. But the cause? The notion is meaningless to me.
References


